AD-A133 145 LOW RATE TRANSMISSION OF VIDEO SIGNALS USING ADAPTIVE DELTA MODULATION(U) CITY UNIV OF NEW YORK RESEARCH FOUNDATION J BARBA ET AL. 15 AUG 83 MDA903-80-C-0476 1/2

UNCLASSIFIED F/G 17/2 NL

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LOW RATE TRANSMISSION OF VIDEO SIGNALS USING ADAPTIVE DELTA MODULATION

Sponsored by the

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

ARPA Order No. 3534

Contract No. MDA 903-80-C-0476

Monitored by Dr. Vinton Cerf

FINAL REPORT

Date of Contract

1 July 80 - 1 July 83

COMMUNICATIONS SYSTEMS LABORATORY

DEPARTMENT OF ELECTRICAL ENGINEERING

THE CITY COLLEGE OF
THE CITY UNIVERSITY of NEW YORK

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

1

LOW RATE TRANSMISSION OF VIDEO SIGNALS USING ADAPTIVE DELTA MODULATION

Sponsored by the

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

ARPA Order No. 3534

Contract No. MDA 903-80-C-0476

Monitored by Dr. Vinton Cerf

FINAL REPORT

Date of Contract:

1 July 80 - 1 July 83

COMMUNICATIONS SYSTEMS LABORATORY

DEPARTMENT OF ELECTRICAL ENGINEERING

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARPA - FINAL REPORT	2. GOVT ACCESSION NO. AD-A133145	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) LOW RATE TRANSMISSION OF VIDEO SIGNALS USING ADAPTIVE DELTA MODULATION		5. TYPE OF REPORT & PERIOD COVERED Final Report 1July 80-1July 83
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) J. Barba, M. Dressler, & D. L. Schilling		8. CONTRACT OR GRANT NUMBER(s) MDA 903-80-C-0476
9. PERFORMING ORGANIZATION NAME AND ADDRESS Research Foundation of CUNY on behalf of The City College, New York, N.Y. 10031		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS ARPA (Code HX 1243) 1400 Wilson Blvd. Arlington, Va. 22209		12. REPORT DATE 8/15/83
		13. NUMBER OF PAGES 145
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) ONR 715 Seventh Avenue New York, N. Y. 10003		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either express or implied, of the Defense Advanced Research Projects Agency or the United States Gov.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

(1) ADAPTIVE DELTA MODULATION
(2) PACKET VOICE
(3) MODEM

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes packet switched voice operation in packet switched networks. The effect of delay in packet reception is considered. A simulation system is described. DSI techniques are discussed as are silence detection algorithms.

The second part of this report describes voice modem, speech waveform identification using delta modulation. The results enable digital transmission at significantly reduced rates.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1

INTRODUCTION

The packet switching concept was originally described by researchers at the Rand Corp. It was a method to be used by the military to achieve a survivable environment for both data and voice transmission.

Today's trends are for people to rely, more and more, on digital computers for support of many of their daily applications. These applications include computer based operations such as electronic mail, airline reservations, information processing, etc.

Information processing of data has become commonplace in our digitally oriented society. Most of the world's information is being stored in digital form and telecommunications is being used as a means to provide access to that information, from both on-site (local) and distant (remote) locations.

Packet switching is a natural extension to methods that are used in computer communications and has received much attention as a method to provide (tele)communications in the computer communications field. The packet switched environment can provide:

1. error free delivery
2. encryption of data for secure communications

3. computer manipulation of data/voice e.g editing, storage, retrieval, etc.
4. code and speed conversion to facilitate and allow for communications between otherwise incompatible locations.

The first three of the advantages listed above are inherent to the digital communications environment. The fourth is the distinct advantage of packet switching.

Data traffic over a network is "bursty" in nature, with long "dead time" intervals between the short transmission segments. The bursty nature of data transmission allows for other services to be provided over the same network. The other services including facsimile, video and audio have come to play a subservient role to data, since the design of packet switched nets has been geared towards data transmission.

Serious attention has thus been focused on packet switched communications. This is because packet switching is the most cost effective technique for utilizing the resources that are available in the bursty data environment, with its high peak to average ratio. Another advantage of packet switched communications is that Packet communications allows for the asynchronous use of the network's transmission resources and for greater sharing of

the transmission capacity than is allowed by FDM or TDM systems.

1.1 PACKET NETWORKING

A packet is defined as a collection of bits whose length varies from a few to thousands of bits long. To these bits a header is added which includes addressing and control information to allow proper routing of the data packet over the network.

Packet networking utilizes digital technology embodied in terminals, computers, modems, multiplexers, error control units and other available devices and techniques.

Figure 1.1.1 shows the structure of a computer communications network. This is the most basic form of a distributed network that is used. There are three basic resources associated with the network [6]:

1. Terminals which are connected to the network either directly (via a concentrator that packetizes, depacketizes, error corrects etc.) or via a HOST computer that is directly connected to the net.

2. HOST computers which perform many tasks, not

only for local users but also as remote processors. The hosts are the information processors of the system. Hosts provide services not only locally and to remote users but also to other hosts

3. The communication subnetwork which consists of trunk lines and switches (ARPANET's IMPs and TIPs). It is thru this subnetwork that the packets are routed.

It is the communications subnetwork that has demonstrated the efficiency of packet communications. With the subnetwork's resources the system's storage, processing and communications capacity must be shared. The ARPANET experience has proved the cost effectiveness of data communications in the packet switched environment along with reliability and throughput.

In complex system design, the first and most critical step is to break down the system into subsystems. It is important that each of the subsystem's functions be correctly defined to minimize the complexity of the interfacing with the subnetwork. The design of the subnetwork can be greatly simplified by providing only one kind of interface, because it is desirable for both voice and data to access a network with minimum standardization. However, public nets have to offer a variety of interfaces, not only for hosts but also for many other devices (e.g.

data terminals, speech terminals). Thus the CCITT/ITU standards committees are of importance to help in establishing various standards in interfacing.

The communications subnetwork should contain all those communication functions which are essential. These include overcoming line errors by retransmittion or bypassing failed parts of the network by rerouting traffic.

1.2 PACKET NETWORK INTERCONNECTION

Packet switched networks make use of various transmission media. These networks have been implemented over public lines; private lines, satellite channels and presently packet switched mobile radio networks are being developed.

Distributed packet switched networks have evolved for use in different environments (using the various transmission media mentioned above) and a need to interconnect them developed. To interconnect the various networks the GATEWAY was introduced. The concept of the gateway is common to all network interconnections.. The interconnection between various nets is of great importance. Since data may be routed from source to sink nod[illegible]gh local networks, public networks international networks and combinations thereof.

The role of the gateway is to terminate the internal

protocols of each net and to provide a common ground across which data from one net to another can pass. A gateway need not be a single monolithic device. The gateway can be a software package at two node switches, connecting the different nets. It may be made up of two parts (one in each net). These parts (gateway halves) may be distinct devices or may be parts of a network switching node. The Gateway can also may be designed to interconnect several different networks. Figure 1.2.1 shows the use of the gateway in the interconnection of various packet switched networks.

1.3 PACKET SWITCHED VOICE

There has been considerable interest in packet switched voice since the feasability of this was demonstrated by researchers at USC/ISI. Packet voice transmission was performed over the ARPANET as early as 1974 and consequently a packet voice protocol was developed.. This protocol led to further experiments both in point-to-point transmission and also in an attempt to teleconference.

Subnetwork facilities sharing of both voice and data can be accomplished because of 40% speech activity observations in two way telephone conversations. Since there is as much as 60% silence in conversations, this area is fertile for research in efficient utilization of the

packet switching environment.

The ISI experiments have encouraged the use of packet networks for real time packet voice applications. In packet switched networks, where speech is to be transmitted, speech much share the subnetworks resources with data (as well as video and facsimile). Data transmission over the packet switched subnetwork is made up of interactive terminals and file transfer of data between various network entities. These both require high reliablity, whereas speech transmission can tolerate some error but needs small delay and constant bandwidth. Small delay in the packet network environment is important to minimize the "dead time" (silence) between speakers.

The use of packetized speech over packet switched networks requires special attention. This is because the mechanism that allows for the efficient handling of bursty traffic causes non-uniform performance in data rates and delay.

1.4 DELAY IN PACKET SWITCHED NETWORKS

The implication for network delay of voice is more subtle than for data, since in addition to response time, dela[illegible] effect users in a subjectively noticable and perhaps objectionable way. The delay factors for packet switched voice may modify the speakers conversational

behavior patterns and change the naturalness of a conversation.

The delay encounered over packet switched networks can be categorized as follows:

1. packetization delay
2. nodal processing delay
3. nodal queueing delay
4. propogation delay

All of these delays add to the random delays that packets experience transversing the communications subnetwork.

The packetization delay results from the time the packet is formed and includes appending the header to the packet along with checksum bits for error control etc.

Nodal processing delay is caused by the various switching nodes over which the packet transversed from the source node to the sink node. This delay includes receiving, processing and outputting to a queue the packets. With an efficient switching node this delay is not significant.

Nodal queueing delay is a function of the rate of packet arrival at an output queue of a switching node and of the capacity of the outgoing transmission line. This delay depends on the previous path the packet transversed

and the statistics of the speech talkspurt. These can be minimized by efficient silence detection algorithms to minimize the packet rate.

The propogation delay of the packets is usually the delay over high speed transmission lines and is usually small and managable.

From experiments performed on the ARPANET [2], using packet voice, it was demonstrated that:

1. Packet assembly time varied from approximately 200msec (for 5.6 pkts/sec) to 50msec (for 20 pkts/sec).
2. minimum transit time was uniformly equal to 250msec for 5.6pkts/sec rate to 20 pkts/sec transmission rate.

In point to point voice communication over any network (circuit or packet switched) two requirements must be met.

1. synchronous output of speech must be maintained to insure good quality speech at the receiver.
2. end-to-end network delays must be small such that the conversation should be natural.

1.5 SCOPE OF THE RESEARCH

This research proposal is concerned with the design and use of a packet voice network simulator (PVNS). Adaptive Delta Modulation (ADM) will be used as the source encoding scheme. DM presents an efficient and inexpensive method to encode speech for toll quality transmission at relatively low bit rates (16Kb/s). The ADM is also a robust device that can tolerate line errors and "leak" these errors off, with speech remaining intelligible, at error rates of less than 10^{-1}.

Previous work [5] has shown that the Song Voice ADM (SVADM) has a higher dynamic range (10-15dB) at all bit rates than the continuosly variable slope DM (CVSD). The SVADM is also the preferred encoding scheme at rates lower than 16Kb/s. Thus the SVADM will be used as the source encoding scheme in this research.

With the use of the PVNS various experiments were performed to demonstrate the usefulness of the PVNS for network optimization. The experiments performed were:

1. Variation of the PVNS' parameters for various silence/speech (s/s) detection algorithms
2. Subjective evaluation of a two-way conversation with fixed delay

3. Network performance as a function of
packet loss and random delay

In the first part of the study two s/s detection schemes (described below) was examined. The PVNS parameters (e.g. $V_\emptyset$, $S_\emptyset$) was varied to subjectively examine the quality of the processed speech. The statistics of the packet rate and packet length will be measured. With the optimized parameters the second part will examine the ease of carrying out a two-way conversation as delay is introduced. The final part will examine the quality of the processed speech as packet loss and random delay are introduced.

Finally, The performance of an ADM as a voiceband data signal identifier will be presented. The feasability of using the digital output of the DM (SVADM) to estimate the spectrum of voiceband modem data waveforms. This spectrum (or correlation) can be used for automatic routing of data/voice bits in a digital network environment.

1.6 SUMMARY OF RESULTS

In this the first section, a brief overview of packet switching is presented. The nature of packet switched systems, their interconnection and delay related issues, which are of importance to packet switched voice, are

discussed.

In chapter 2 the first part of dissertation research is presented. The issue of packet switched voice, currently being implemented over telephone lines and numerous local area networks, is discussed. The design of a real-time packet voice network simulator is described along with the associated hardware and software aspects.

Chapter 3 addresses the importance of silence/speech detection. This section briefly describes various analog and digital schemes that exploit the nature of silence periods in conversational speech, to increase channel utilization.

Chapter 4 shows experimental results obtained from the real-time packet voice simulator. Packet rate statistics and transmitted packet size statistics are plotted along with with the subjective results of various threshold values used. This is done for one of the silence/speech detection algorithms, where the optimal parameters are chosen.

In chapter 5 the second part of the dissertation is presented. It is shown that various synchronous modems can be discriminated one from the other, using the autocorrelation function of the digital output of the delta modulator. This is done for the purpose of directly demodulating the analog voiceband modem waveform. Since by direct demodulation of the analog waveform a tremendous

savings in bandwidth can be achieved over PCM encoding of the waveform.

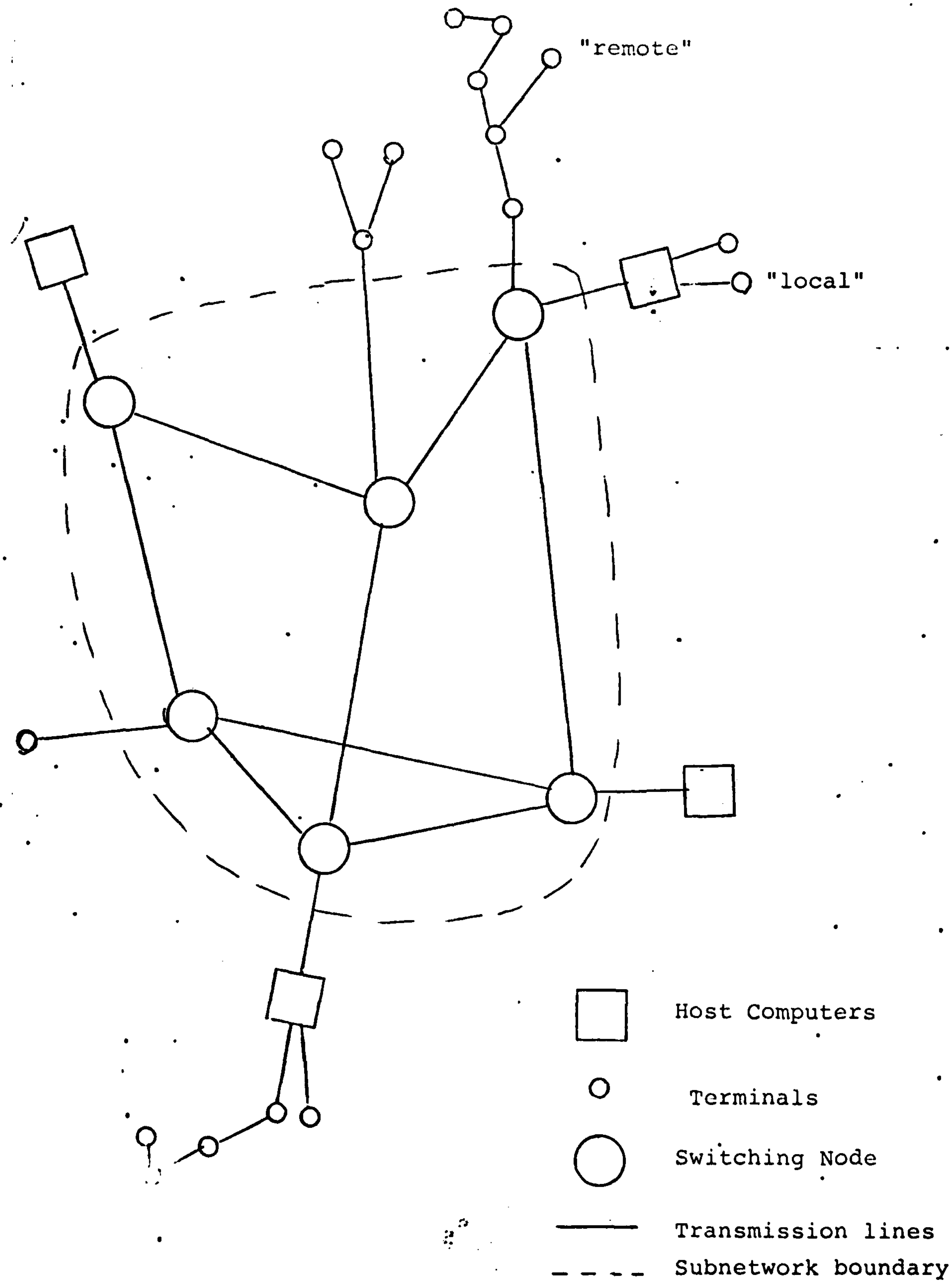


Figure 1.1.1 Computer Network structure

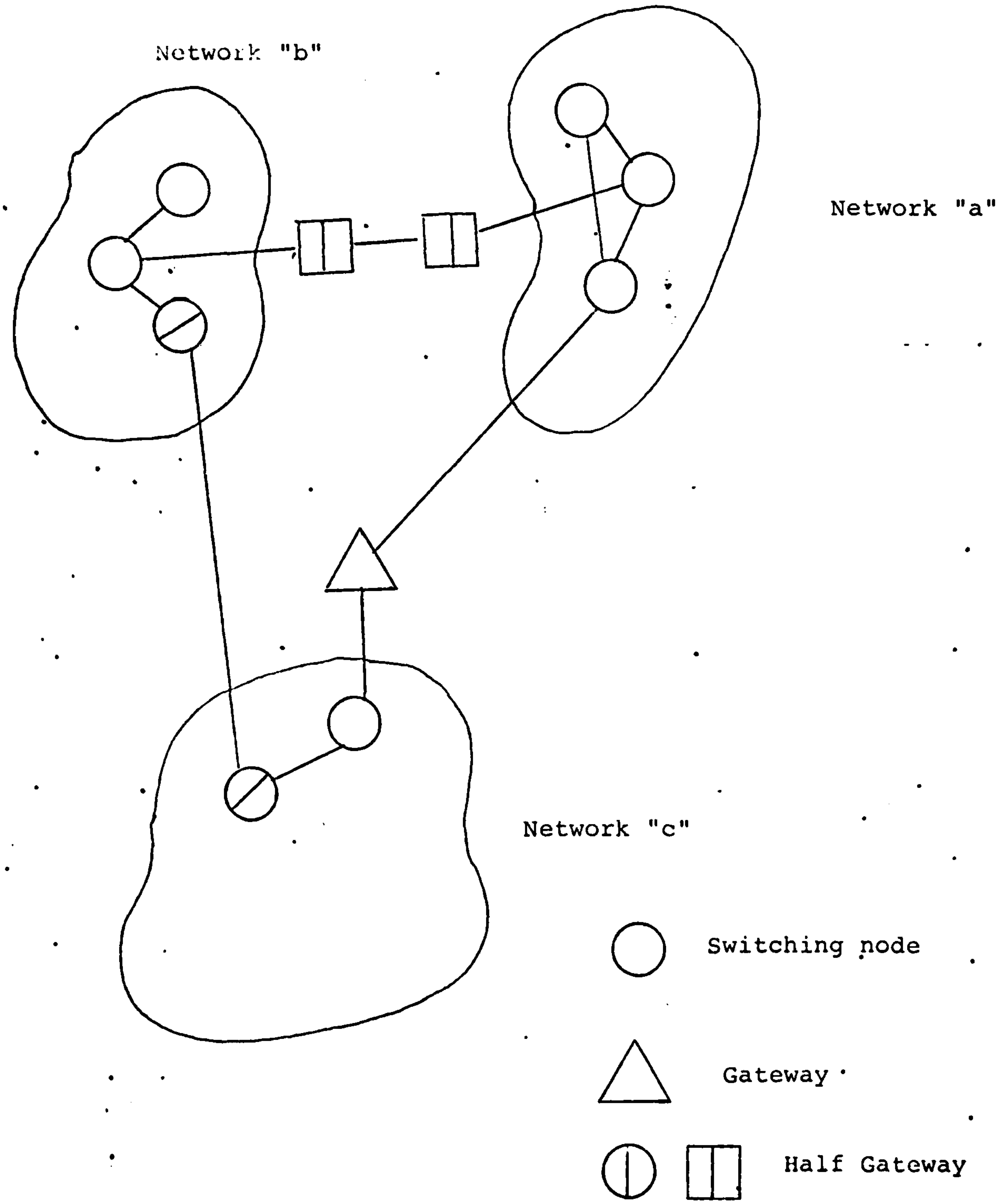


Figure 1.2.1 Interconnection of Networks

Chapter 2

SPEECH TRANSMISSION SYSTEMS

Packet communications is, as described earlier, a scheme which is suited for handling the bursty nature of data traffic. Speech, which is characterized by burstiness and a high peak to average ratio, is an atractive candidate for the packetization process and would increase the transmitting channel's utilization.

The price paid for higher channel utilization is packet overhead (e.g. headers) along with processing and queueing delays, which add to the variable delay experienced by the packets in transversing a network. The most important issue in a voice transmission environment is speech quality.

Maintaining speech quality and the naturalness of a conversation entails having continuity of phrases and words. The control of average end-to-end delay, which contributes to the continuity of speech, is of great importance to overcoming the physiological effect of delay.

The consideration of packet switching, for speech communications, is to take advantage of the low bit rates available for encoding speech and the silent periods inherent in conversational speech. Potential savings in transmission, with respect to 64Kbps circuit switched voice

could be greatly increased. For example, using 50% speech activity and a 16Kbps encoding algorithm channel utilization can be increased by a factor of 8 [(1/.5)X (64/16)].

2.1 ISSUES IN PACKET SWITCHED VOICE

The transmission of speech over a network requires continuity of phrases and talkspurts, to maintain the quality of real-time conversational speech. Control of end-to-end delays due to the variability of packet transmission is also of importance in maintaining the naturalness of speech. Therefore, issues such as routing and packet protocols must be addressed.

Packet voice protocols in a store-and-forward environment revolve around two possibilities, Virtual circuit and datagram service. Virtual circuit service is a scheme whereby the packets, to be transmitted, travel over a preassigned path and consequently arrive in the transmitted order at the receiving end. Although this protocol sounds attractive it has its limitations, in that a path must be set up initially and can lead to congestion at some nodes. The advantage include composite packet possibilities and preservation of the transmitted sequence. Datagram service is a scheme whereby packets travel thru the network over any allowable path. This scheme, although

not preserving the message sequence and the added overhead of addressing (additional header information) can be used at high transmission rates.

Routing strategies are also of importance and are related to the protocols used. The routing strategies can be either fixed or adaptive and a combination of both can be supported by a network. Table 2.1.1 and 2.1.2 list advantages of the packet voice protocols and of the routing strategies mentioned above.

The issue of optimal packet size for speech transmission is also of importance to insure network efficiency. A study performed [26] to optimize the packet size used tradeoff between packetization delay (which decreases with packet size) and network delay (increases with packet size). Values of packet size from 300 bits (for low speed vocoders) to 700 bits (for PCM encoding) were found to be optimal and results in packet rates of 15-30 pkts/sec. This rate is very high and from work performed on large store-and-forward networks would result in extremely high delays[2]. Table 2.1.3 lists the effect that packet size has on the performance of a packet switched network.

2.2 [illegible]MENTS OF SPEECH PACKETIZATION

Speech transmission over any communications network

(circuit switched as well as packet switched) is in reality a point to point transmission, the generalized form of a packet switched network (fig 1.1.1) can be simplified as shown in figure 2.2.1. This is the generalized packet (in our case packet voice) switching network system model with which we will be working.

At the transmitting end an analog speech waveform is encoded into a binary sequence bit stream, since a conversation is continuous (silent periods are also encoded), this bit stream is then fed into a packetizer. The packetizer examines the bit stream that is generated and efficiently detects the start and the end of speech talkspurts. At the instance of speech detection the process of packetization begins. The speech "talkspurt" is a fundamental concept of the packetization process and was defined by Brady [11] as follows:

1) Speech power exceeding a threshold value for a time greater than 15msec.
2) Seperation from neighboring talkspurts by a time interval known as a hangover.

The talkspurt definition is illustrated in figure 2.2.2. This figure clearly shows that if the hangover period is less than a specific value the periods of talker activity constitute a single talkspurt. The 15msec

threshold value is a means to distinguish speech activity from impulsive background noise. As discussed above the hangover is used to bridge short silence periods within a talkspurt. Hangover values of 160-240msec are typical. The speech power threshold that is used for defining the speech talkspurt should depend on the hangover value (e.g. -40 dB for 200msec (relative value of 0 dB is where overload occurs)).

At the start of the packetization process the packet network's system time (which must be established in common by all speaker locations) is added to the header of the packet, for later use at the receiving end. The timestamping of the packets is of importance to preserve the continuity of the speech. Figure 2.2.3 a and b show the packetization process of a speech talkspurt. When silence is detected the packetization process ends only to be restarted by further speech activity.

When packets are formed, they are presented to the packet subnetwork. Within this subnet the packets are routed from node to node via whatever routing algorithm that has been established on that network. Since the path any one packet travels over the subnet is not necessarily the same as another packet, the packets suffer random delay and do not necessarily arrive at the receiver in the same order as they have been transmitted in, this is shown in figure 2.2.3c. Due to the random arrival of the packets

at the receiver they must be correctly reassembled for playback.

To reconstruct, in the correct order, the transmitted packets it is necessary that the receiver check the present system time with the packet formation time (stored in the header). The packet is then inserted at the correct location, relative to the other packets, within the output buffer, as shown in figure 2.2.3d. If the packet arrives after it should have been output it is discarded (lost packet) as illustrated in figure 2.2.3e, which shows one packet of the talkspurt that is discarded. The packets being in the correct sequence at the receiver buffer are then depacketized.

The depacketization process involves serially outputting the bit stream through a decoder and playback is achieved at the receiving end.

2.3 PACKET VOICE NETWORK SIMULATOR (PVNS)

The packet voice network simulator effectively simulates a packet switched network environment. The block diagram of the PVNS is shown in figure 2.3.1.

The packet network functions, e.g. packetizer, communications subnetwork and depacketizer are all performed by a PDP 11/34 minicomputer. These functions which basically constitute the packet network are the

software aspects associated with the simulator.

The hardware aspect of the PVNS is basically the data connecting equipment. The hardware portion of the PVNS uses a serial/parallel (transmitting end) converter attached to an encoder at the input. The output is connected to a parallel/serial converter (receiving end) which is in turn connected to a decoder. for this simulator both the input and the output are multiplexed between two channels.

The simulator allows for the study of packet voice in real time. The subjective study of a two way conversation to examine the aspects of delay on a conversation is also made possible. The statistics of a one way conversation along with various qualitative and quantitative aspects associated with the packetization/depacketization process are also feasible.

2.3.1 HARDWARE CONFIGURATION

A PDP 11/34 minicomputer was used along with a DR11-K parallel input/output interface, to connect the computer to the external devices. The specification of the control device (data connecting equipment) used to interface the ext[illegible] devices, the data terminating equipment (consisting of a pair of encoder/decoder) to the computer (using TTL logic ICs) is as follows:

1) 16 bit parallel input (output) to (from) the computer for each channel.
2) 16 bit parallel to (from) serial conversion at the decoder (encoder).

Figure 2.3.1.1 shows how the speech data input to the PDP 11/34 is accomplished. An analog speech waveform is encoded by two independent encoders, which are clocked by an external source. The sampled data is then serially input to 16 bit shift registers simultaneously (one for each encoder(channel)). Corresponding to the 16th bit, data is latched into a multiplexer which presents it to the DR-11K parallel interface. Upon storing the first data word into the main memory the DR-11K issues a signal which causes the other channel's data to be loaded into the multiplexer and a similar process of storing the data word into the PDP 11/34 is performed. The reading of the data from the DR-11K input bufffer to the PDP 11/34 cache memory is performed by the software portion of the PVNS.

The output of the data to the data connecting equipment is shown in figure 2.3.1.2. The data to be output is transferred from the PDP 11/34 cache memory into the DR-11K output buffer. From the output buffer the data is transferred into a multiplexer, which in turn transfers the data to shift registers performing the parallel/serial

conversion process. After the data for the first channel is entered into the output multiplexer, control circuitry signals the DR-11K to output the other channels data word. A similar process is then performed on the second channel to output the data serially. The data which is present in the output parllel/serial shift registers is then clocked out at the rate of the external clock to a pair of decoders.

Continuous bit streams are thus generated for the decoders of both channels. It is noted that data is read-out of the computer after every read-in operation. The speed of operation of the PVNS is limited to the speed of writing in and out of the PDP 11/34 (e.g. machine instruction cycle time) and the rate of the external clock.

2.3.2 SOFTWARE CONFIGURATION

The PVNS, as discussed previously simulates a packet switched conmmunications subnetwork for two independent voice channels. The simulation program consists of approximately 600 lines of machine language instructions, which allows for the real-time operation of the simulator. The general flow diagram of the simulator is shown in figure 2.3.2.1.

The first step taken by the program is to initialize the data area that is used by the simulator. Figure

2.3.2.2 shows how the data area, of one of the two channels is apportioned. Memory locations Ø thru 262_8 are used for control, data and pointers. Some of these include the channel's status register, the input and output words, various pointers etc. Locations 264_8 thru 362_8 are used as a shift word register, These store the latest 64 bytes of data. From location 1000_8 onwards there are 256 blocks of packet control words (headers associated with the packets) and data buffer area (packet buffers). The data area dedicated to a channel comprises 4K bytes (256 blocks along with pointers, status register etc.) and 32K bytes of packet buffer space used for each voice channel. This makes a total of 36K bytes of memory used for each channel of the PVNS. The packets that are created by the initialization process are formed in a chain like manner, as shown in figure 2.3.2.3. These packet control words (PCW) are aranged in a manner that one packet points to the next and the previous packets, along with a pointer to the data buffer area associated with that particular PCW. The buffer data area stores the actual information bits that are input from the external hardware thru the DR-11K interface.

After the initialization process is completed the simulation starts as follows: Two words are read into and two words are read out from the PDP 11/34 main memory, as described in the previous section. At this point the base

address register is set to point to data area dedicated to the first channel. After the input/output operations are completed, the processing of the input data is performed sequentially for each of the two channels followed by the output processing. When the processing for both channels is completed, two words are input and output to and from the encoder and decoder pairs and the process continues.

Figures 2.3.2.4 and 2.3.2.5 show the operation of the input and output operations of the PVNS. The processing sequence starts with the input processing portion of the simulator. The process operates on the data as follows:

1) Voice detection (if in silence mode)
2) Silence detection (if in voice mode)
3) Allocation of the packet buffer
4) Random delay time generation (with constant added delay, if desired)
5) Insertion of packet buffer into the proper location of the output packet buffer pool (chain).
6) Continuation of Packetization, for as long as the simulator is in voice mode

The outputting of data is done sequentially for both channels after the inputting of data as follows:

7) Comparison of packet output time with present system time and decision to output

8) Outputting of either words from output packet buffer area or silence patterns (if in silence mode).

To perform these tasks we divide the packet buffer area for each channel into three different packet buffer chains (refer to figure 2.3.2.3). These are the idle chain, input chain and output chain. At initialization all the packets created are assigned to what we call the idle packet chain. When speech is detected by the simulator, a packet is acquired from the idle buffer chain and placed within the input packet chain. The previously stored words (from the shift word register) and the newly arrived word are stored in the packet buffer area. The packets that are placed to the input chain are time stamped with the present system time along with a random and possibly fixed delay and are then inserted in increasing order for transmission to the output buffer chain.

When a new packet is created and its output time is assigned (stamped), the packet should be inserted into the proper location within the output buffer chain. This is done by a searching operation performed on the output cha[illegible] Process number 4 (listed above) requires considerable processing time. For example, the number of

output packets which exist in the output buffer chain can be greater than 40 in some instances. Since the period of time that is possible for each input/output 'cycle' is limited (input to the computer is continuous) for real time operation. 'Cycle' is defined as the unit of time from the input of a channel to the next input of the same channel. The cycle is therefore a function of the external clock rate and all of the system time values are normalized to this unit. Processes nos. 4 and 5 which are performed at the time of new packet creation, are time divided into several sequential tasks. Each of these tasks is executed within the single word (16 bits) processing cycle. If N cycles of the search operation are performed to locate, for insertion of the packet into, the proper location in the output buffer chain. N+3 cycles in total are needed to complete the processing.

2.3.3 RANDOM NUMBER GENERATOR

The random number generator (RNG) is of great importance to the PVNS. It is used for generating random time delays in simulating the network delay experienced by the packets transversing the net and as part of the packet los[illegible]nism.

The numbers generated by the RNG should in reality be called pseudo-random, because they are obtained from a

determined calculation. However, the numbers generated pass many of the statistical tests for randomness. Since these numbers are obtained from a known calculation the sequence will eventually repeat. The numbers that a process will generate before repeating is called the length of its period. Some processes degenerate repeating the same number (usually 0) while other processes will produce very short sequences.

There are two schemes generally used for generating random numbers. The numbers are generated sequentially and derived in a simple way from numbers preceeding it. In one way the number is obtained by a process of addition, in another way by multiplication.

The most commonly used method to generate RNs is the multiplicative and is as follows:

$$X_{n+1} = CX_n \quad \text{(modulo word length of the machine)}$$

The multiplicative RNG process can easily be seen. Each number of the sequence determines the following number, etc. The main problem involves the selection of C (multiplier) and X_0 (seed). The RN process is described by Ha[illegible] in [27] states that C as well as X_0 should be odd numbers.

All odd numbers C can be written in the following

ways:

$$8t\pm1 \text{ and } 8t\pm3$$

for some t [27].

It is proved that C=8t-3 will have a period of $2^{k-2}-1$, where k is the number of bits used in the calculation by the machine.

Figure 2.3.3.1 lists the program that was used to generate the random numbers. Values of 37 were chosen for both the seed and the multiplier. This number was found by using t=5 in the formula above and would suggest a sequence of length equal to $2^{(15-2)}-1$ $(2^{13}-1)$ = 8191 numbers would be possible.

Lines 3 to 6 of the RNG program establish the maximum numerical value of the RNG, which is limited to the value stored in BL. Lines 7 to 9 multiply the seed [and subsequent random numbers (not limited by the value BL)] by the multiplier. Lines 10 and 11 takes the value, which results from the multiplication, and by operating on that number limits the value of the RNs to generate a maximum value of $2^{BL}-1$ (e.g. BL=7 corresponds to a maximum value of $2^7-1=127$).

Tests were performed on the RNG program to find the period of the sequence and the distribution of the RN generated. Figure 2.3.3.2 shows the measured period of the

RN sequence using C and X_0=37 for various BL. This measurement was done by comparing the first few generated RNs with subsequently generated values. Except for BL=2 values for BL greater than 4 produced the expected length of the sequence equal to 8191.

Figure 2.3.3.3 shows the period of the RN sequence with C=37 and several values of X_0. It is seen that for odd values of X_0 the maximum period is produced and even values produce a sequence less than maximum. Figure 2.3.3.4 shows that if the multiplier is changed and the seed is both even and odd the sequence length is very short (and at times degenerative). The distribution of the RNG was also performed and it was found to have a uniform distribution (as expected).

Assuming an external clock rate of 16Kbps, 1024 bit packet length and continuous speech (continuous packetization) we are interested in how much time the RNG's sequence lasts. Since a RN is added to the header at packet creation, a sequence of length equal to 8191 implies that the RN does not repeat for 524 seconds (8 minutes and 44 seconds). For conversational speech, where packetization does not occur continuously the sequence could last for as much as 2.5 times longer.

Virtual call	Datagram
1 Call setup packets	1 addressing overhead in every packet
2 call storage time	2 Potential for high speed transmission
3 call option processing	3 message sequence not preserved
4 composite packet possibility	

Table 2.1.1 Packet Voice Protocols

Fixed Path	Adaptive Routing
1 Abbreviated headers	1 Overhead due to routing table updating
2 Delay subject to local conjestion	2 Robust against failure
3 call lost if path fails	3 Reduction in delay by local conjestion avoidance
	4 Must be constrained for heavy load

Table 2.1 .2 Routing Strategies

Long Packets	Short Packets
1 Good Overhead efficiency	1 Low overhead efficiency
2 longer queue delay	2 Shorter queue delay
3 Shorter processing delay	3 longer processing delay
4 less processing complexity	4 greater processing complexity.

Table 2.1.3 Packet size tradeoff

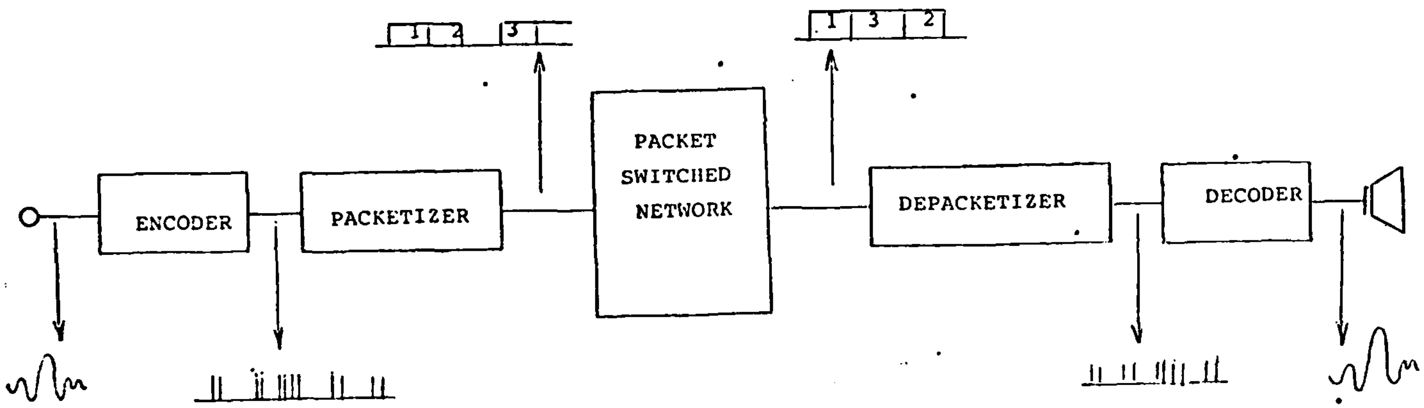


Figure 2.2.1 Packet voice transmission system

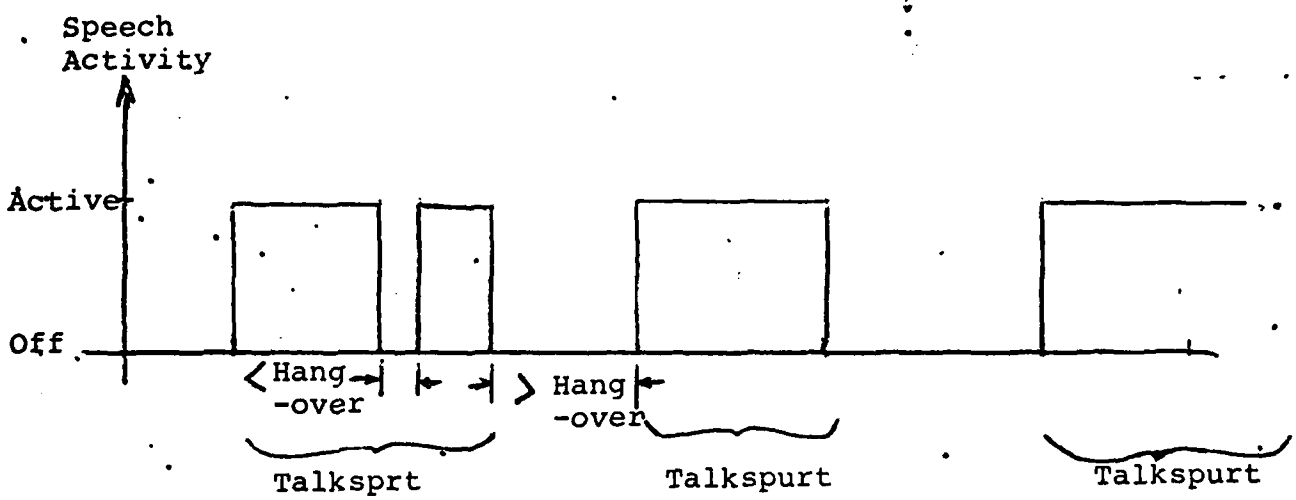


Figure 2.2.2 Definition of a Talkspurt

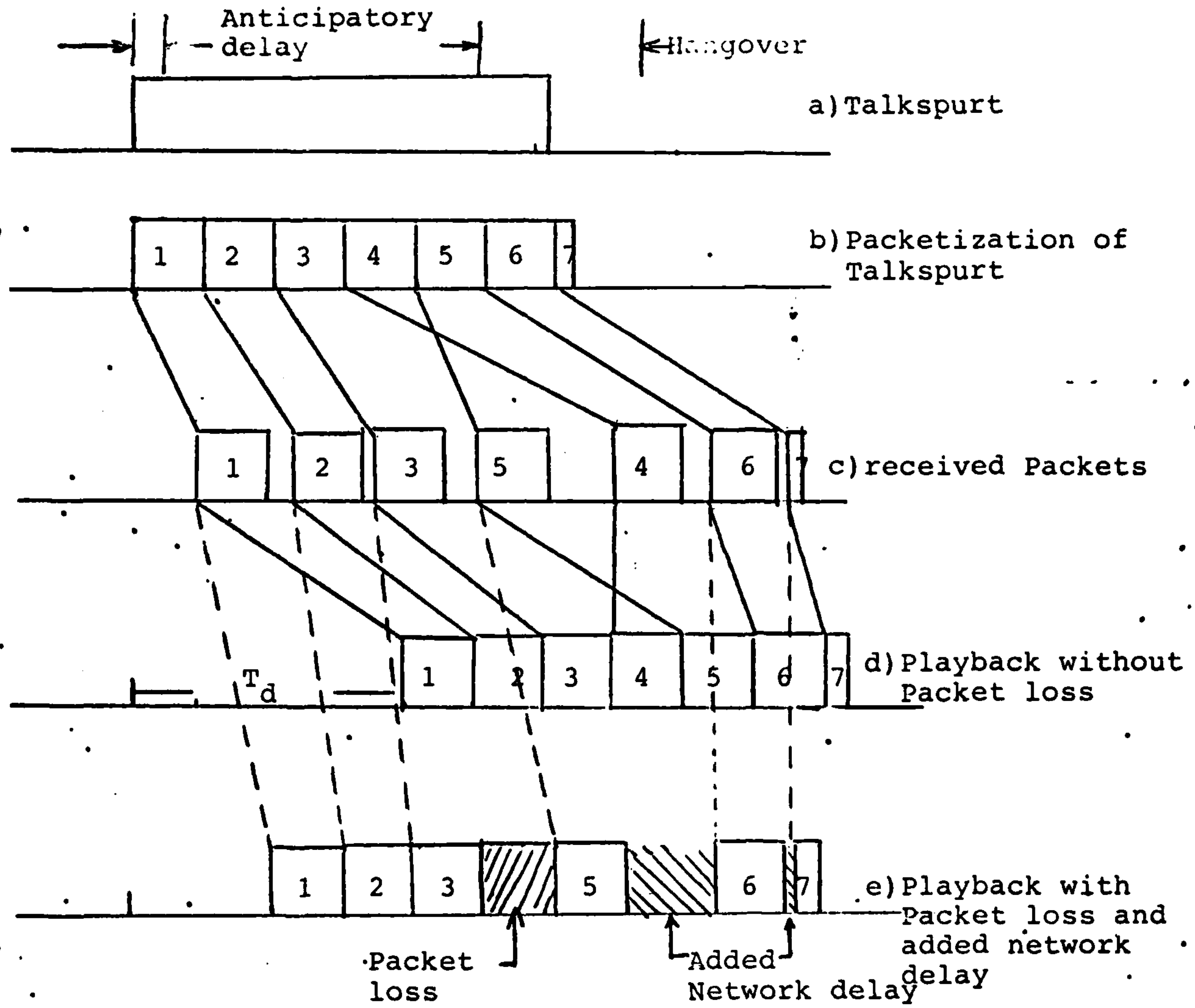


Figure 2.2.3 Packet transmission of a Talkspurt

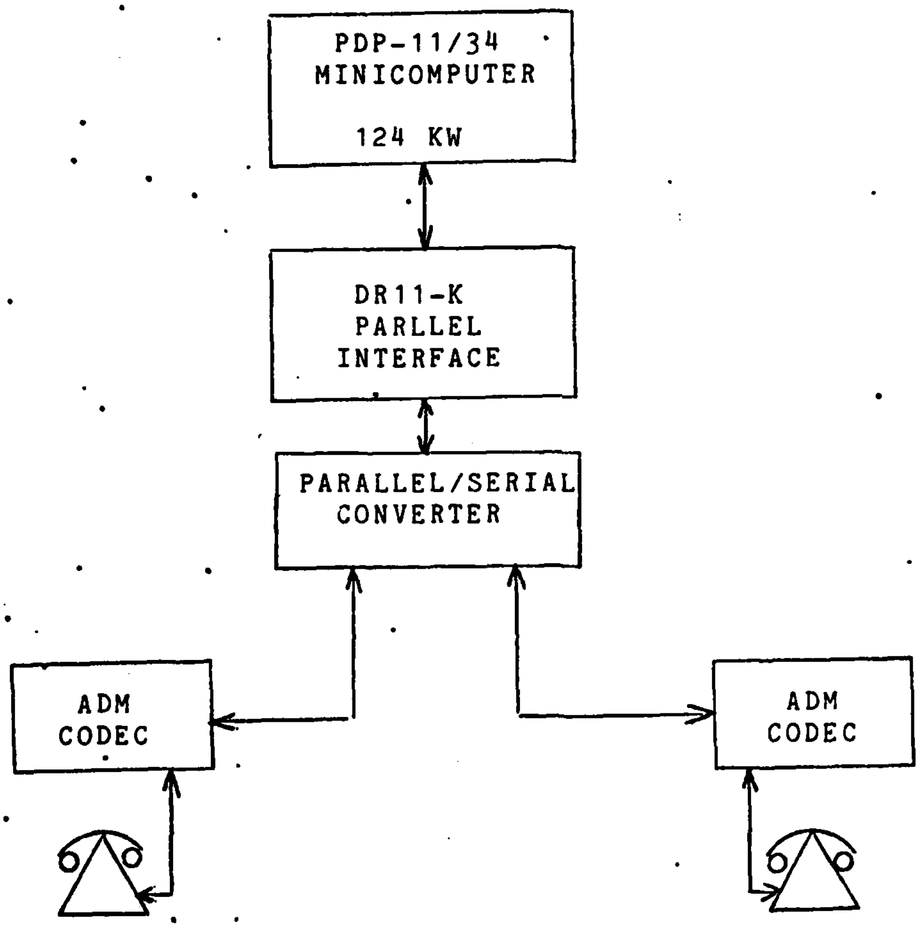


Figure 2.3.1 Packet Voice Network Simulator

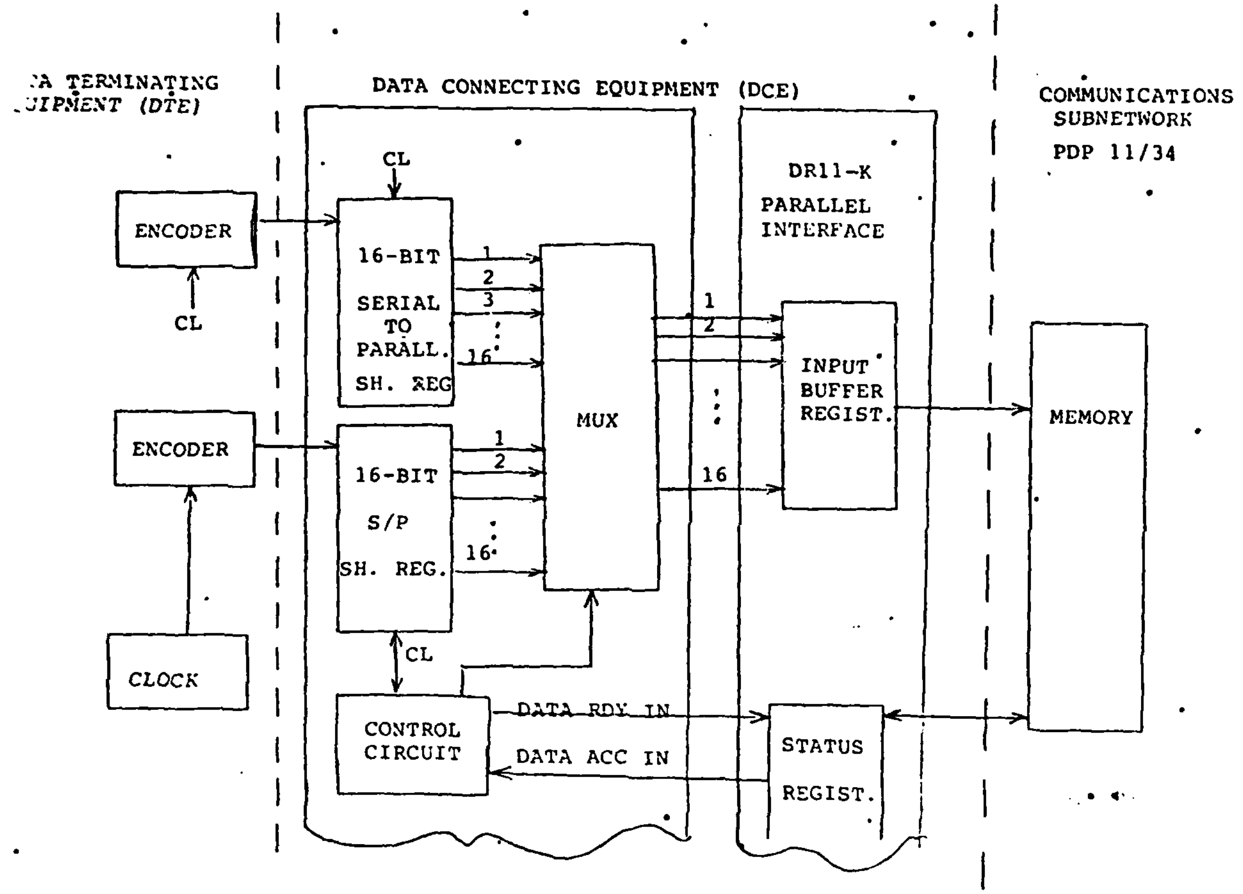


Figure 2.3.1.1 Configuration for data input to PVNS

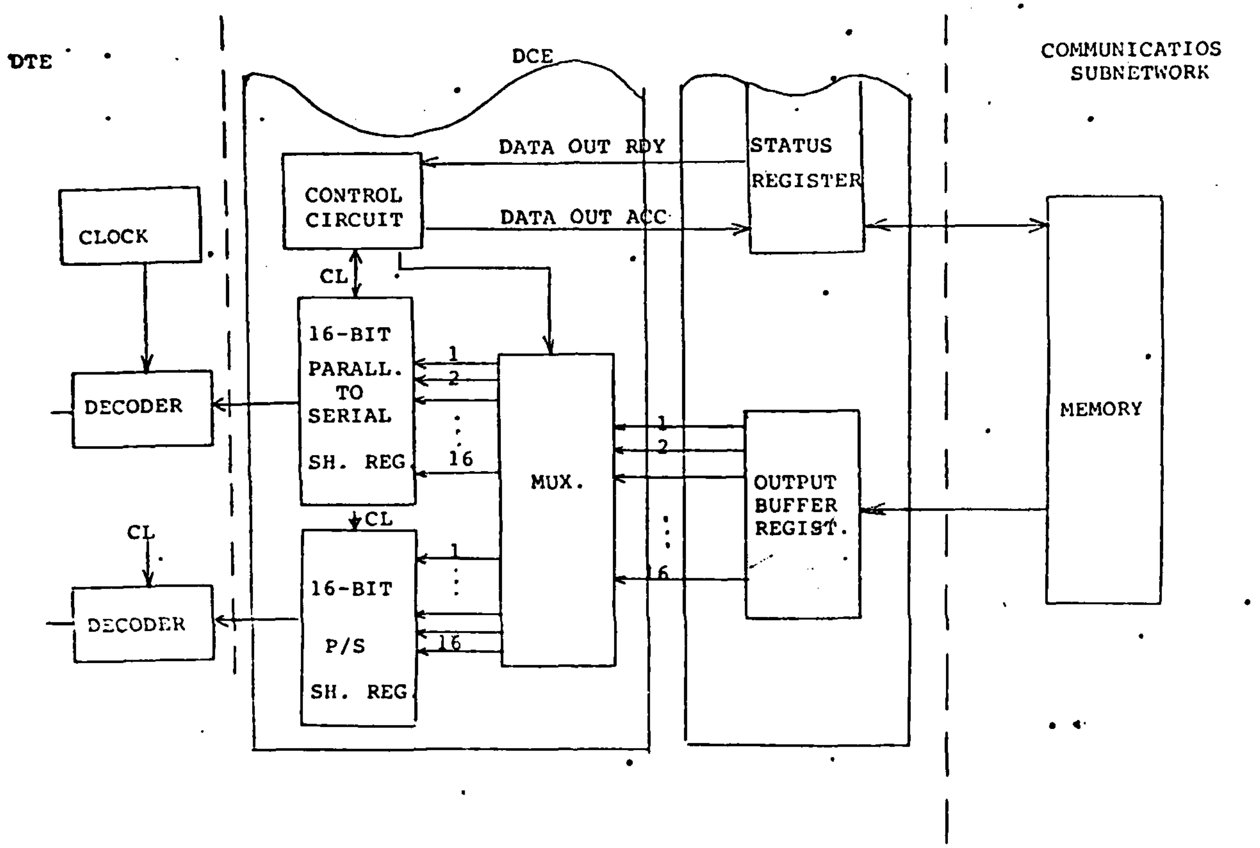


Figure 2.3.1.2 Configuration for data output from PVNS

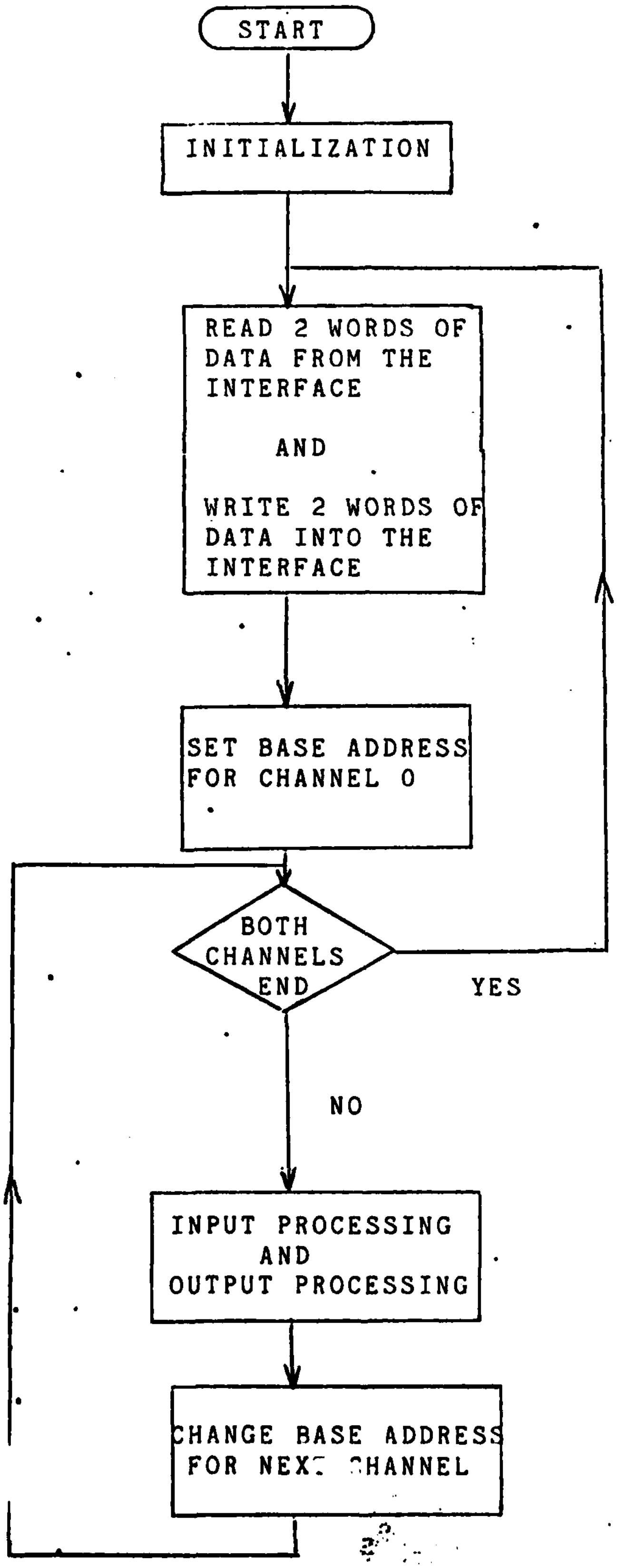


Figure 2.3.2.1 General flow chart of simulator

MEMORY LOCATION	
0	*CHANNEL STATUS REGISTER*
2	INPUT DATA
4	PREVIOUS INPUT DATA
6	OUTPUT DATA
10_8	IDLE CHAIN POINTER
12	INPUT CHAIN POINTER
14	OUTPUT CHAIN POINTER
.	
.	
264_8	SHIFT WORD REGISTER (64 BYTES)
.	
.	
362_8	
.	
1000_8	PACKET CONTROL WORD #0 — 7 WORDS
.	
1016_8	PACKET CONTROL WORD #1
.	
.	256 BLOCKS TOTAL PER CHANNEL
10000_8	PACKET BUFFER # 0
MAXIMUM LENGTH OF PKT. (200_8 B)	
10200_8	PACKET BUFFER # 1
.	

Figure 2.3.2.2 Memory map of control and packet data area for channel 0

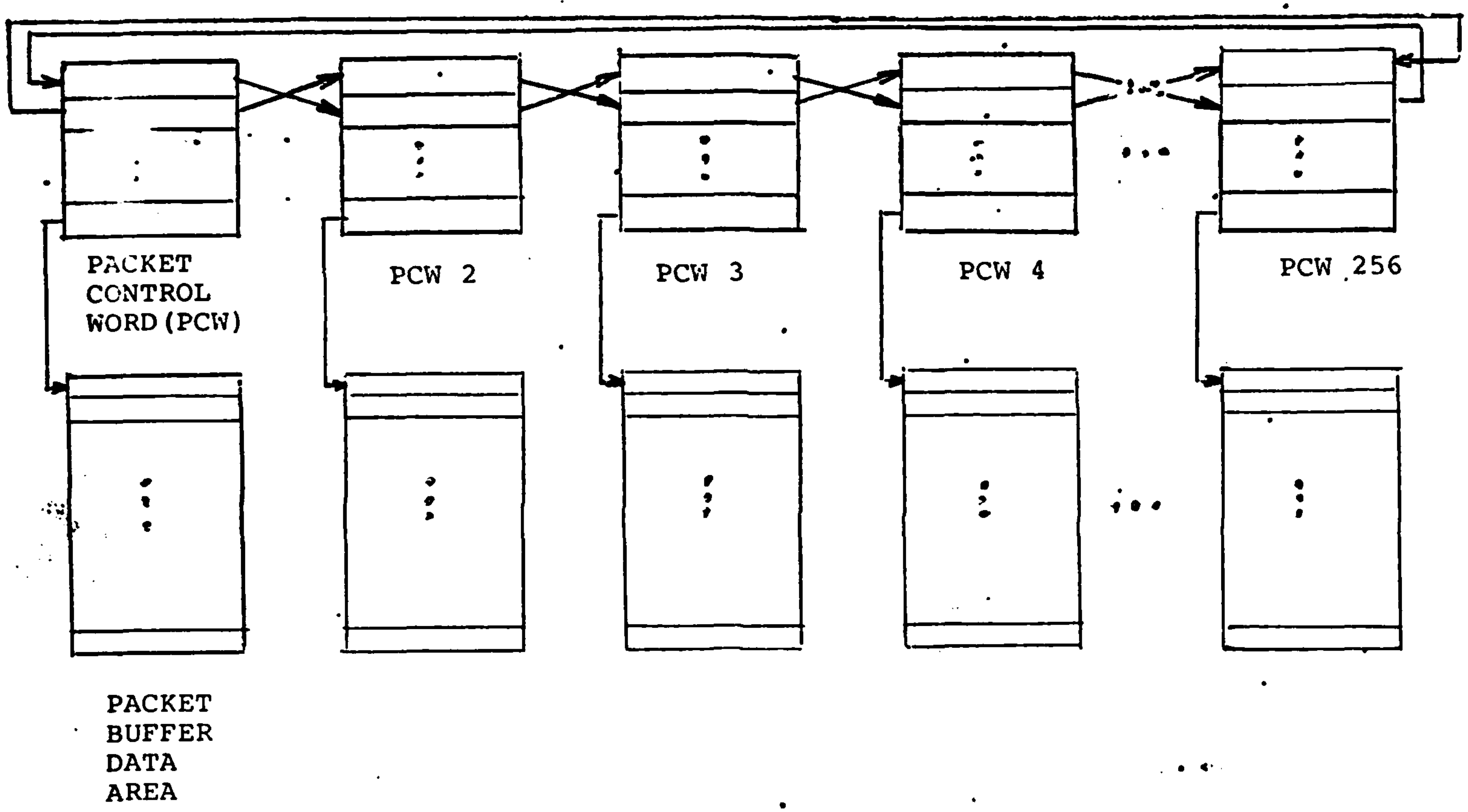


Figure 2.3.2.3 Packet chaining at program initialization

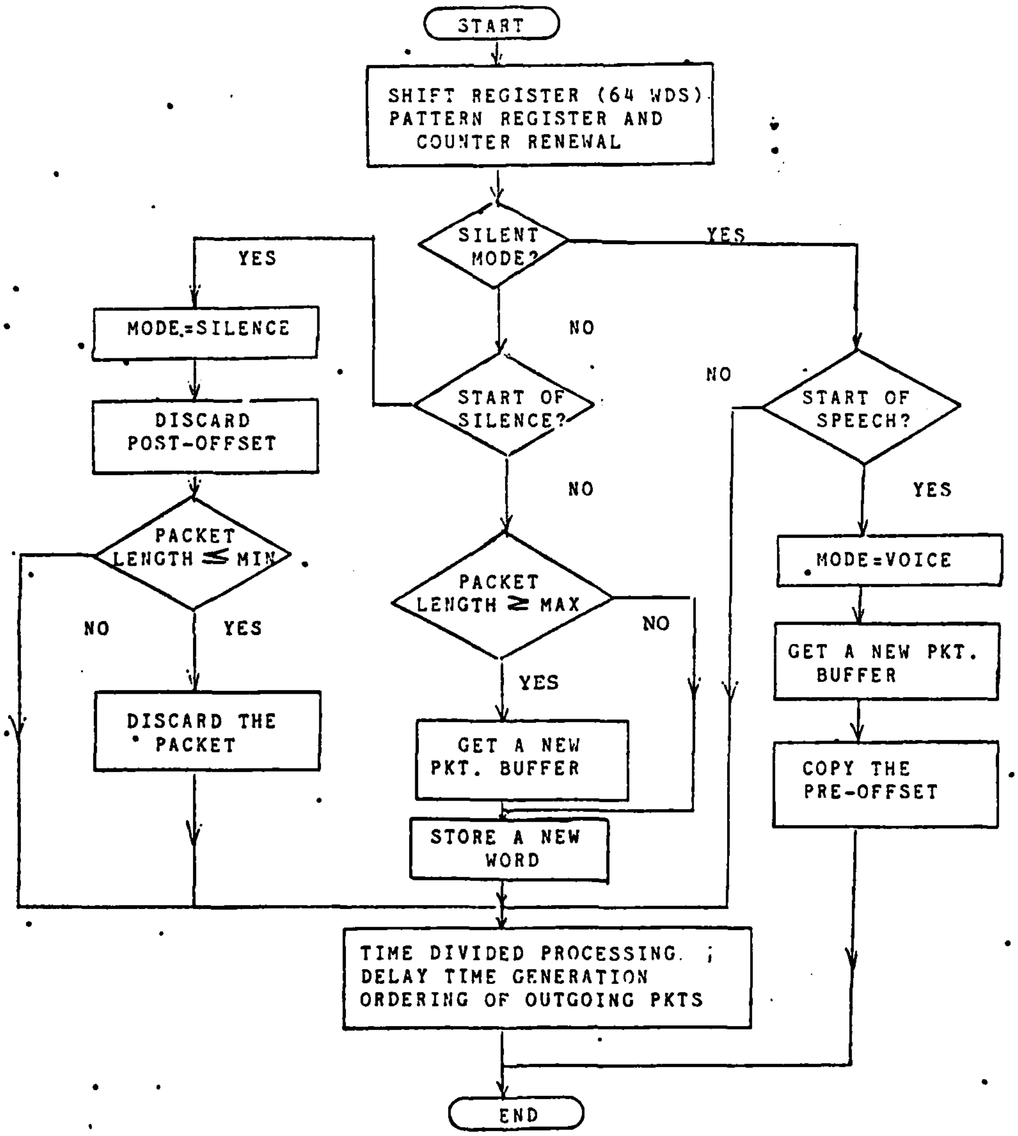
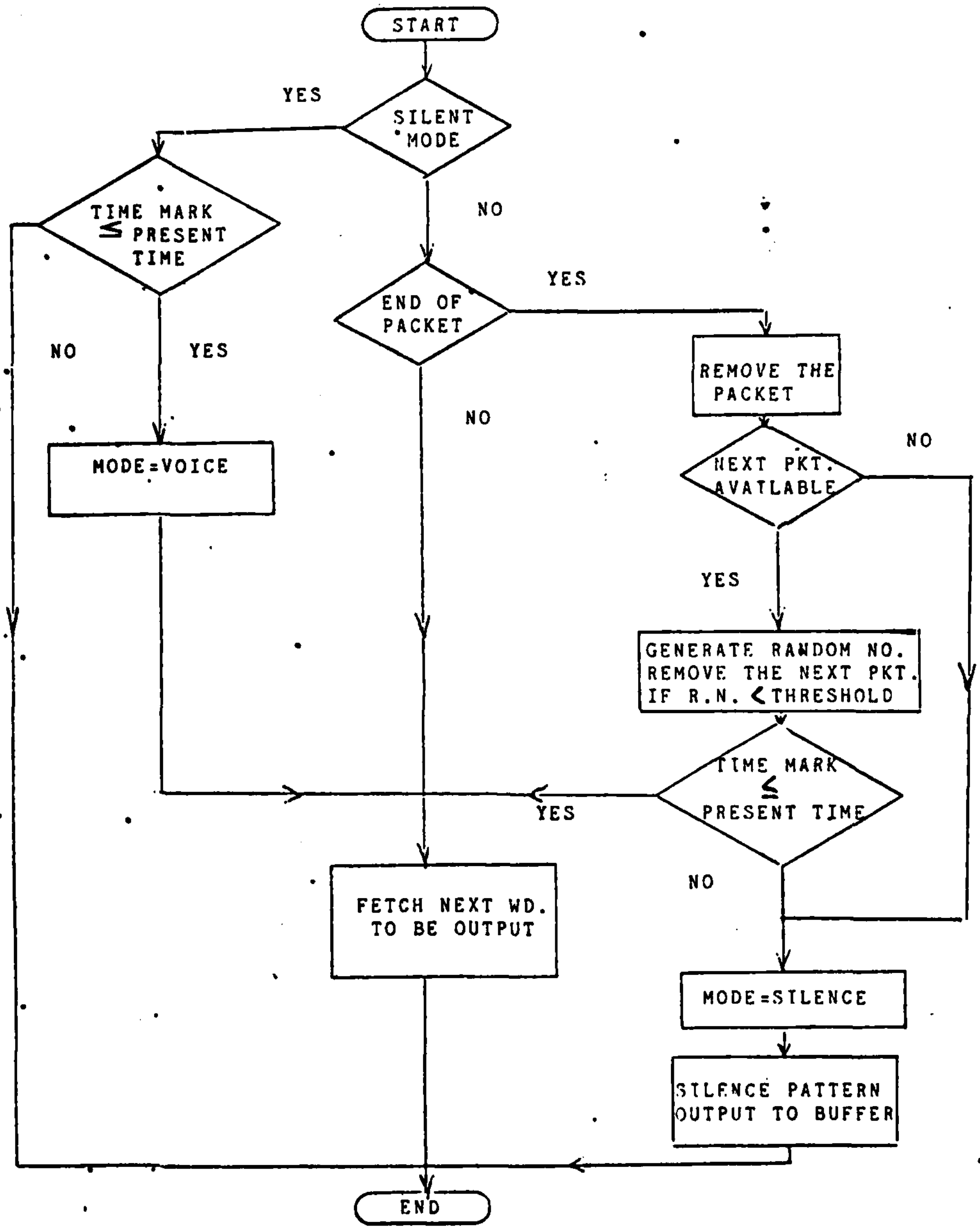


Fig .3.2.4 Flowchart for input processing



F... 2.3.2.5 Flowchart for output processing

```
            MOV #37., @#X      ;X is seed
            MOV #37., @#C      ;C is multiplier
            MOV @#BL,R4
            MOV #15.,R5
            SUB R5,R4
            MOV R4,@#SH
      RAN:  MOV @#C,R0
            MUL @#X,R0
            MOV R1,@#X
            BIC #100000,R1
            ASH @#SH,R1
```

Figure 2.3.3.1 Program to generate Random Numbers

(Length of RN sequence)
No. of R.N.'s compared in sequence

B_L	4	5	6
2	(192	498	1317
4	8196	8197	8198
5	8196	8197	8198
7	8196	8197	8198
9	8196	8197	8198
12	8196	8197	8198)

Multiplier=37
seed=37

Figure 2.3.3.2 Measured period of RNG

B_L	Length of R.N. Sequence (No. of RN's compared in sequence =6)			
	seed= 10	22	45	65
2	4102	965	4083	5389
7	4102	4102	8198	8198
12	4102	4102	8198	8198

Multiplier=37

Figure ? 3.3.3 Length of RN sequence for various seeds and constant multiplier

SEED	MULTIPLIER	(B_L=9) LENGTH OF R.N. SEQUENCE
37	16	3
37	20	7
37	30	14
37	50	14
20	16	3
20	20	6
20	30	12
20	50	12

Figure 2.3.3.4 Length of RN sequence for various multipliers and constant seed values

Chapter 3

SILENCE/SPEECH (S/S) DETECTION FOR PACKET SWITCHED SYSTEMS

In using speech over packet switched networks, it is important to preserve the quality and ease of a conversation.

The quality of a conversation is dependent on preserving the continuity of the speech talkspurts. The ease, on the other hand, is dependent on the end-to-end delay experienced by the packets transmitted over the subnetwork. A long delay experienced can lead to detrimental physiological effects on a conversation.

The effect on the communication subnetwork of voice transmission is different than that of data. When a conversation takes place, over a communications subnetwork, there is a requirement for constant channel bandwidth. This is because a conversation is continuous and the resources of the subnet must remain available for the duration. Therefore, it may be necessary to give priority to voice packets over data packets. With voice priority end-to-end delay constraints can be maintained, even in the presence of data packets.

It was determined by Brady [11] that there is greater than 50% silence in a conversation, and as much as 60-65%

one way. This characteristic has been, and is presently being exploited to improve channel utilization.

3.1 TIME ASSIGNED SPEECH INTERPOLATION (TASI)

TASI, initially, was an analog method introduced in the late 1950's by Bell Labs. In the Speech Interpolation (SI) Scheme, speech talkspurts are interpolated within silent periods. SI is one method to improve channel utilization amongst the network's users, SI has evolved over time and is still being improved and changed, since it was introduced.

The process of speech interpolation becomes more efficient when the number of channels increases. For example, when two conversations are to be interpolated over a single channel, It is possible that, due to contention for the channel much of the conversation will be lost for until the talkspurt is terminated it has control of the channel and could "feeze-out" attempts by other talkspurts.

Freezout consists principally of short clips of the initial portion of a talkspurt ranging from zero to several hundred miliseconds in length. Periods of freezouts greater that 50msec were measured to be perceptively damaging to the quality of speech. The problem of speech loss (freezout) was due mainly to the speech detection scheme that was used. Contention for the channel was

another problem that was addressed and corrected in the early 1960's.

3.2 DIGITAL SPEECH INTERPOLATION (DSI) [12],[14]-[16]

In digital speech interplation (DSI) conventional PCM/TDM data (at 64Kbps) is input to the DSI system. The data coming from the PCM/TDM channels is used by a transmit assignment processor. The processor using a digital voice detector assigns the n-ary input trunks to a m-ary TDM/TASI channels frame, which is then transmitted to a receiver where the process to output the data is reversed.

The advent of digital technology corrected the defficiencies of the analog TASI systems. The problem of freezout was addressed by methods of variable quality coding. During periods of overload on the system, the least significant bit of all the transmission slots is reapportioned to augment the transmission capabilities of the system. Other advantages of DSI over analog TASI include the fact that digital voice detectors perform much better than their analog counterparts. Also more precise switching of digital speech samples among the channel slots is possible.

Two forms of DSI that have been implemented are the Speech Predictive Encoded Communications (SPEC) and the ADPCM-TASI systems. The SPEC system used predictive coding

and was implemented over a satellite channel. ADPCM-TASI systems used a fixed rate channel and bits, over the channel, are allocated such that the coder stays in synchronization with the channel rate.

One of the principal merits of the SPEC system is total avoidance of TASI's clipping problem. The SPEC system uses an improved voice detection scheme that improved noise immunity of the speech detection and enhanced the switching characteristics on speech bursts, leading to an improvement in the operation of DSI systems.

In the ADPCM-TASI scheme proposed by McPherson [16] bits are allocated evenly among all the active speech channels. Another method proposed by Yatsuzuka [14] uses a voiced/unvoiced detector along with silence detection.

3.2.1 BUFFERED SPEECH INTERPOLATION (BSI) [13],[15]

Buffered speech interpolation or variable rate coding TASI is the most recent advance in the SI techniques. It is an improvement on the DSI techniques described above. In this technique talkspurt delay is used to mitigate the freezout impairment.

Buffers are used, in these schemes to store excess data and then to transmit them over the network. The buffers effectively decouple the coders from the channel.

Elder and O'Neill [15], using a PCM system stored the

data in a buffer. In this scheme some users experienced no delay while others experienced variable delay. Cox and Crochiere [13] proposed a method of variable rate coding that introduce the same delay among all the users. This method is based upon buffer fullness, not on the channel rate of the previously mentioned scheme. Speech activity and an increase/decrease in coder rates were also used to prevent overflow/underflow of the buffer.

The BSI techniques were initially being used in a multiserver, statistical-TDM, teletrafficing mode rather than in a single server, store and forward packet voice mode. Cox' scheme, however, also addressed the use of buffers in packet voiced systems not only SI system.

3.3 SILENCE/SPEECH (S/S) DETECTION ALGORITHMS FOR PVNS SYSTEM

To efficiently allocate the system's resources and to preserve the continuity of the speech, effective silence detection must be used. In our experiments Delta Modulation (DM) will be used as the source encoding method of the speech waveforms. It is the special characteristic of the DM that enables us to use them for s/s detection and is now addressed.

3.3.1 SVADM FOR S/S DETECTION

The response of the SVADM to a constant input is shown in figure 3.3.1.1. It is seen that for the constant input the Delta modulator's output oscillates in a ...11001100... fashion. For periods of silence, this is the pattern that is generated by the DM.

3.3.2 CVSD FOR S/S DETECTION

As in the SVADM algorithm, the CVSD algorithm also has an oscillating pattern for a constant input. This pattern is a ...101010... pattern, as shown in figure 3.3.2.1.

3.4 SILENCE DETECTION ALGORITHMS

Two methods to detect s/s in a conversation were implemented as part of the PVNS. The encoded input to the computer's serial/parallel interface is stored into a shift word register (SWR), whose length is Lmax bytes. The value Lmax is varied with the bit rate of the Delta Modulators. The SWR's length is determined by previous work, in speech detection, to equal approximately 30msec of input data.

3.4.1 16 BIT SILENCE DETECTION SCHEME

Figure 3.4.1.1 shows the flow diagram used for s/s detection for this scheme. A 16 bit word is entered and the presence of any silence pattern is checked. If the word corresponds to any of the possible silence patterns, shown in figures 3.4.1.2 and 3.4.1.3, the s/s counter is incremented (with a maximum value equal to the length of the SWR L_{max}). For a word that does not match the pattern the counter is decremented. To determine the mode (silence/speech) of the converstion we use threshold values $V_{\emptyset}$ (speech) and $S_{\emptyset}$ (silence), as shown in figure 3.4.1.4. Now that the value of the s/s counter has been updated we check if the conversation was previously in silence mode. If in silence, is the s/s counter less than $V_{\emptyset}$. A positive response indicates that we have speech and packetization begins. If not, we remain in silence mode. If the previous mode was speech, is the counter greater than $S_{\emptyset}$. A positive response ends the packetization process and a negative response keeps us in the speech mode (packetization continues).

When we initiate the packetization process, It is important to include all the speech stored in the SWR. Thus, a few of the previously stored words from the SWR are included at the head of the packet. This Pre-offset enables us to include all of the initial speech segment (avoids clipping).

At the end of packetization (onset of silence) a few

words are deleted from the tail of the last transmitted packet of the talkspurt. This post-offset enables us to remove some of the silence from the end of the last transmitted packet.

Figure 3.4.1.5 shows the response of this scheme to a typical portion of a talkspurt. The numbers within each sampling interval indicate the value of the s/s counter. V_0 equal to 6 and S_0 equal to 6 were chosen, and L_{max} was set equal to 10.

3.4.2 WORD BY WORD (PATTERN MONITOR WORD) S/S DETECTION SCHEME

This scheme is illustrated in figures 3.4.2.1 thru 3.4.2.3. The method for entering and storing words in the computer is essentially the same as discussed above for the 16 bit s/s detection scheme.

If the present word corresponds to a silence (or speech) pattern, then is the previous word also silent (speech). If the response is affirmative the pattern monitor counter is incremented, otherwise it is reset to zero. This scheme although using 16 bits of data is similar to a bit-by-bit s/s detection scheme. For silence (speech) detection, if the counter is greater than $S_0 \cdot (V_0)$ we have silence (speech), otherwise we are in the same mode as we were in previously.

Figure 3.4.2.4 shows the response of this scheme to the same talkspurt as shown for the first scheme. The rest of the discussion (e.g. with regards to pre and post-offset) also applies here. Values of V_0 equal to 4 and S_0 equal to 4 were chosen for this example. L_{max} is set equal to 10.

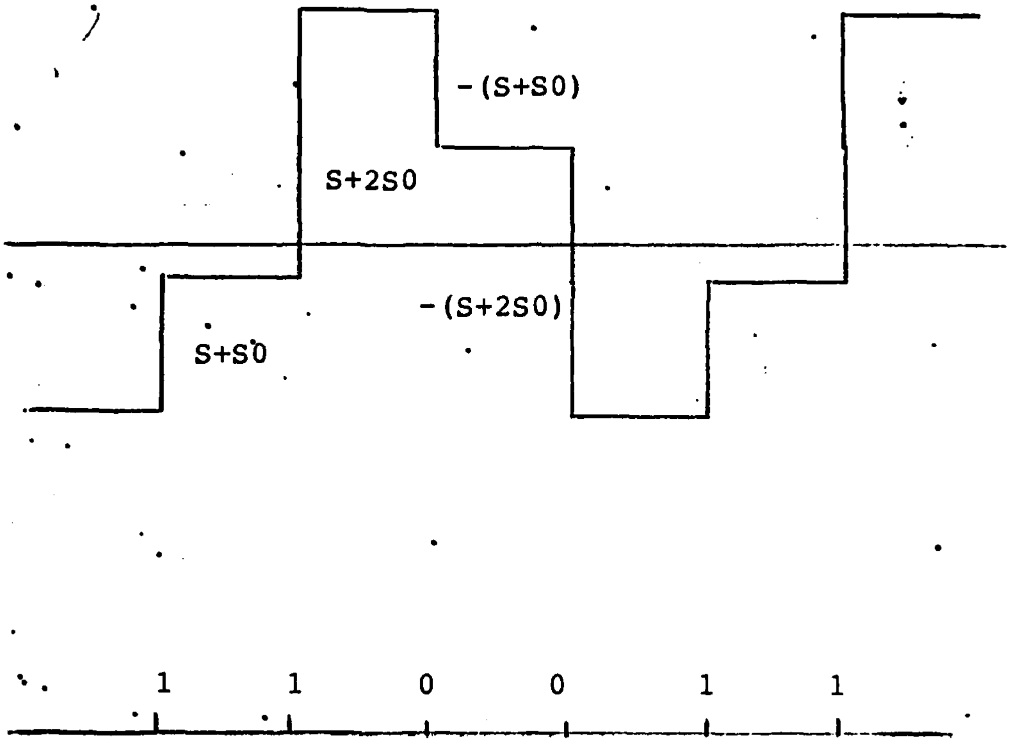


Figure [illegible].1.1 Response of SVADM to constant input

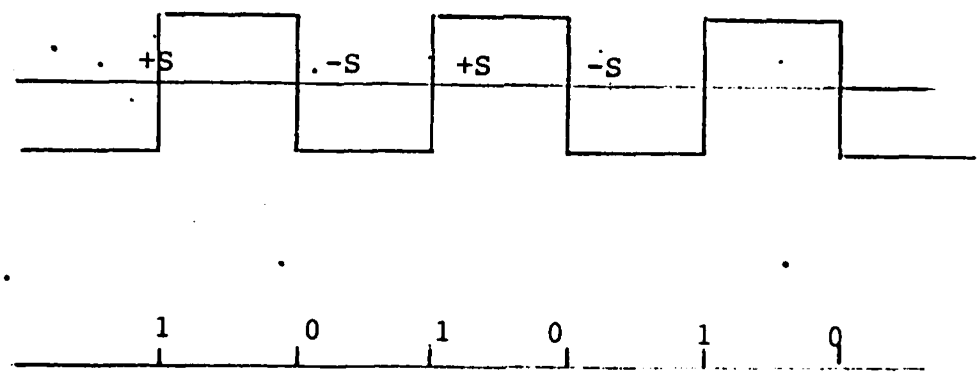


Figure 3.3.2.1 Response of CVSD to constant input

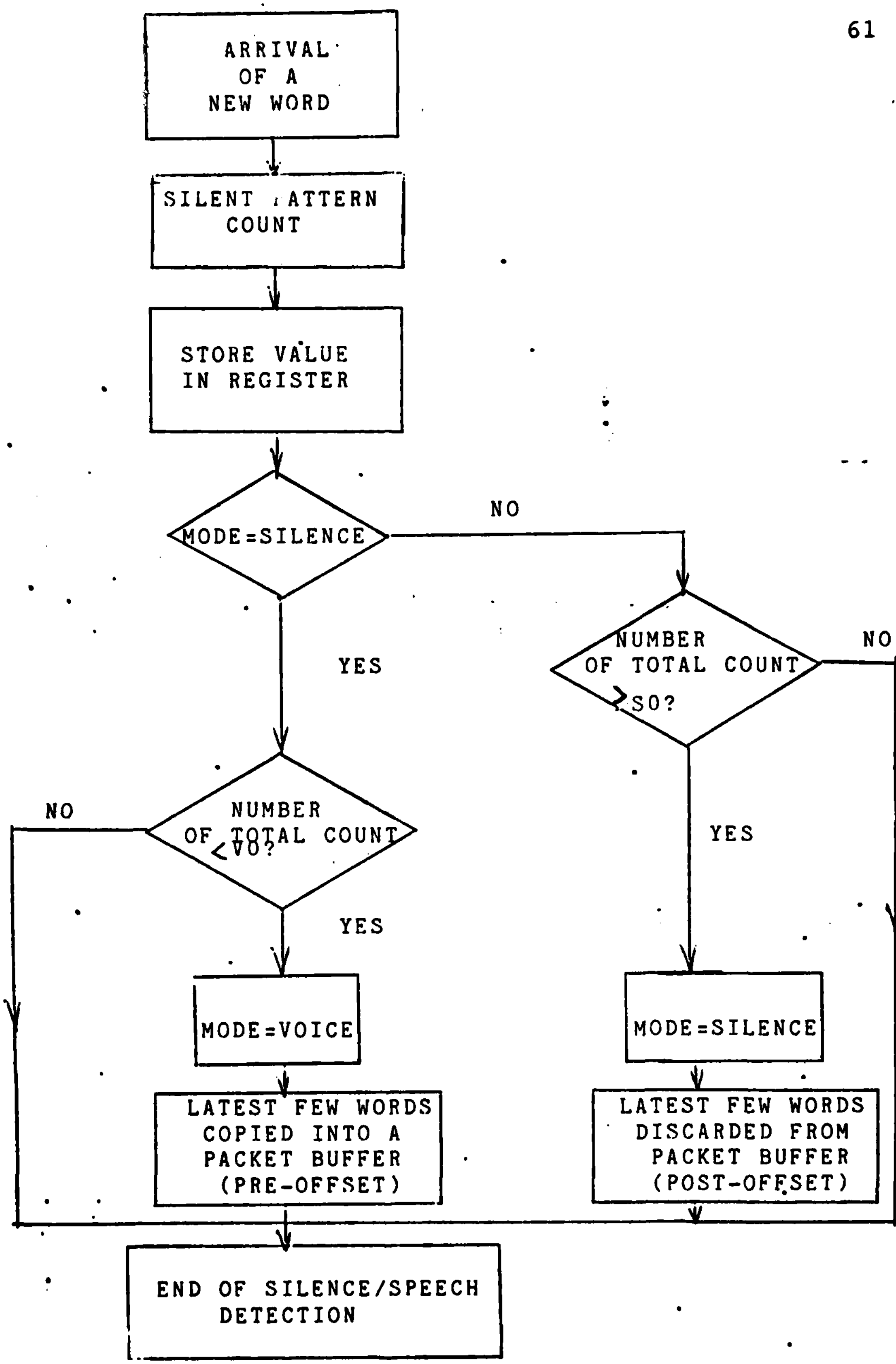


Figure 3.4.1.1 Silence/Speech detection scheme

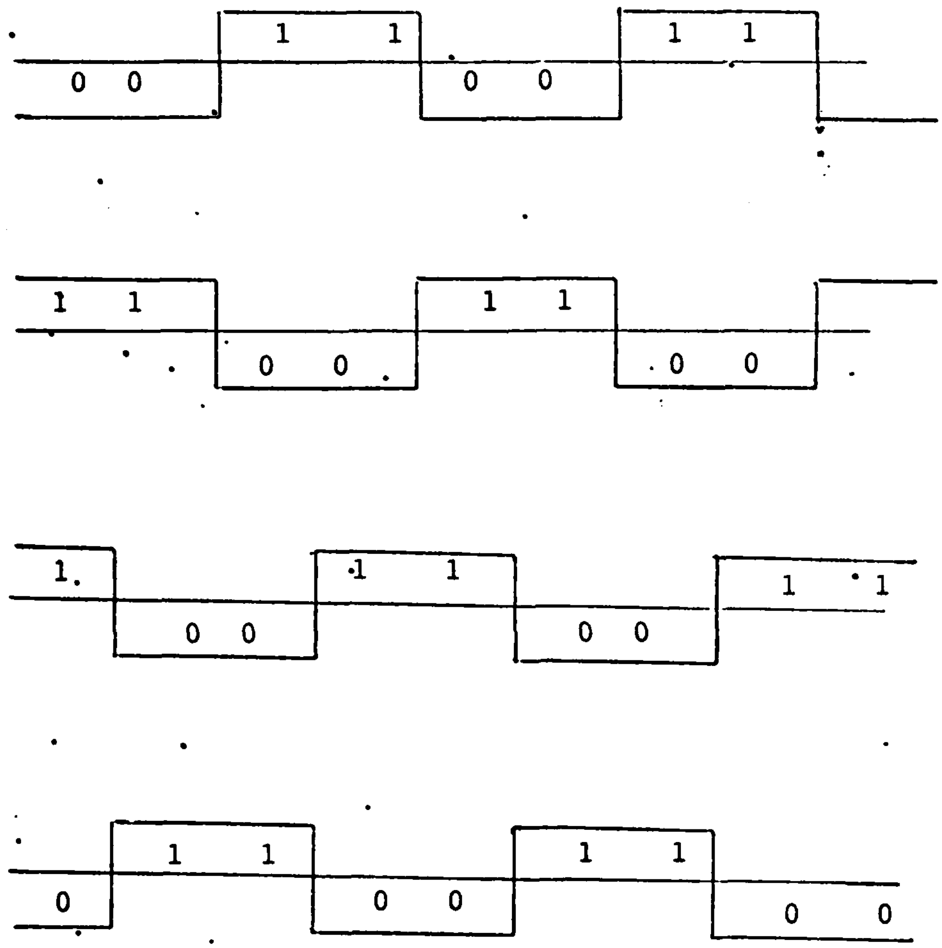


Figure 3.4.1.2 Possible silence patterns for SVADM

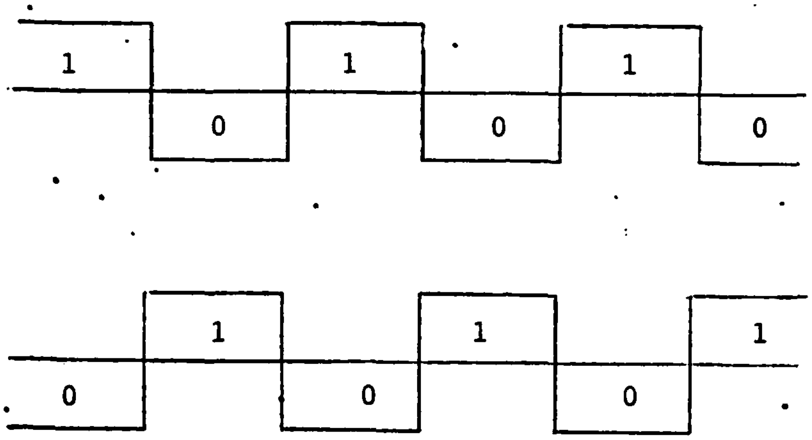


Figure 3.4.1.3 Possible silence patterns for CVSD

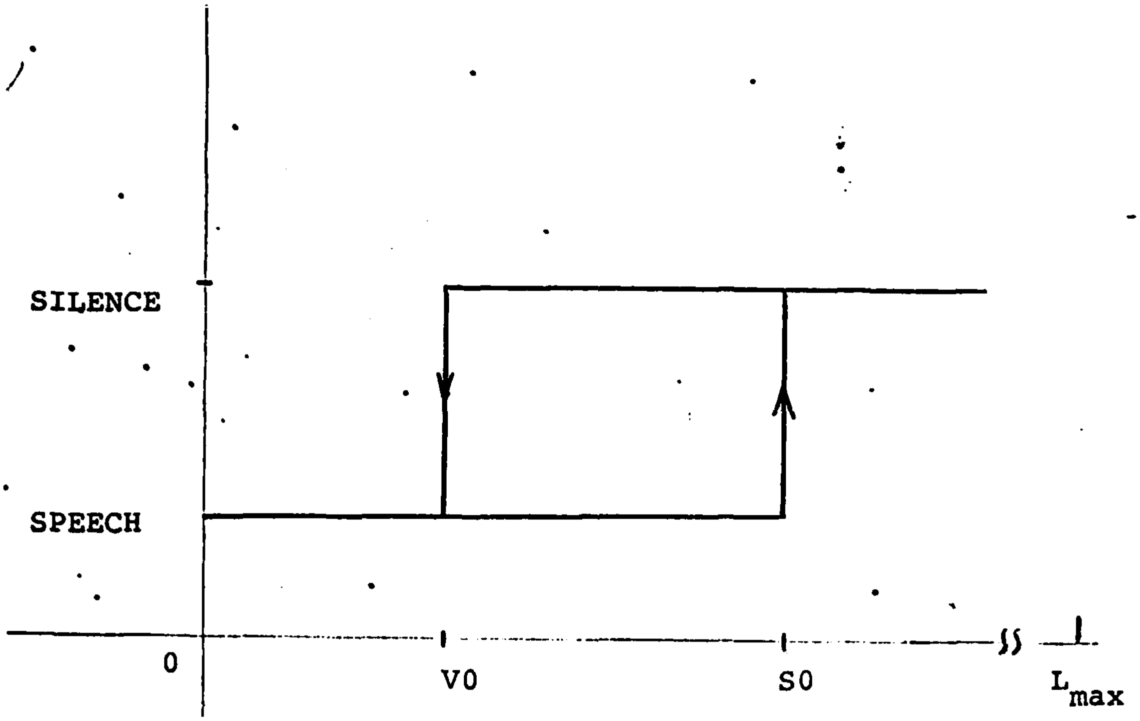


Fig. .4.1.4 Hysterisis mode s/s threshold detection

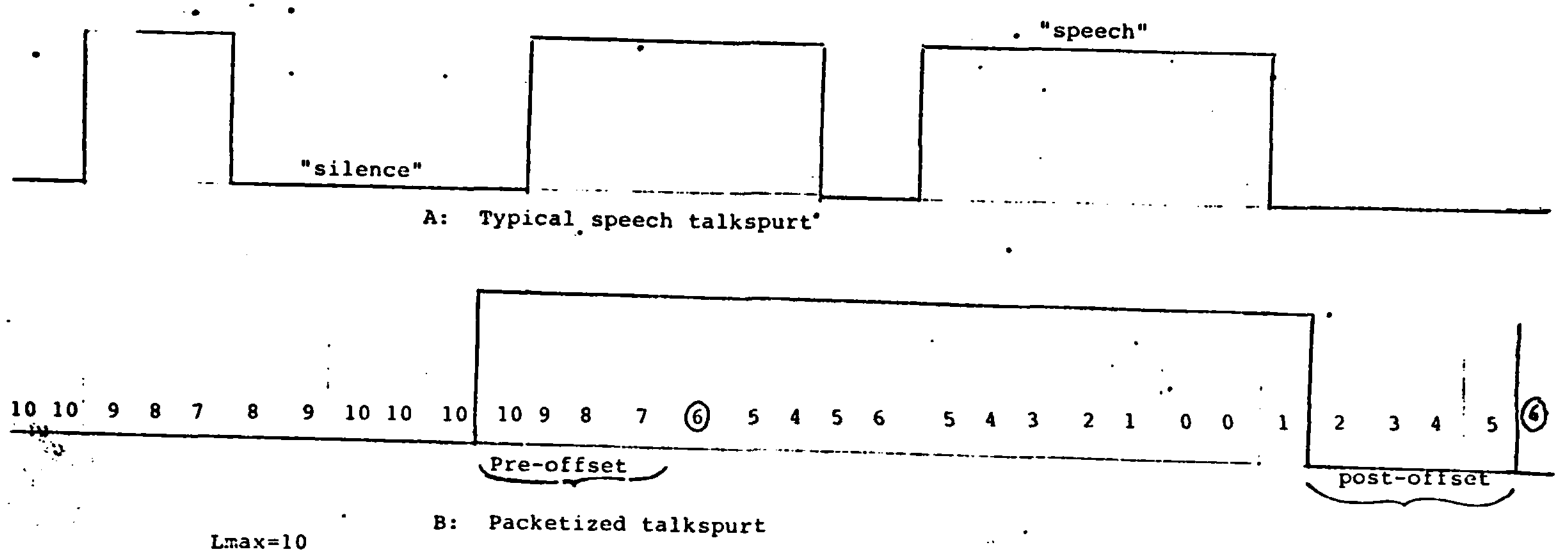


Figure 3.4.1.5. Packetization of 16 bit s/s detection scheme of a talkspurt

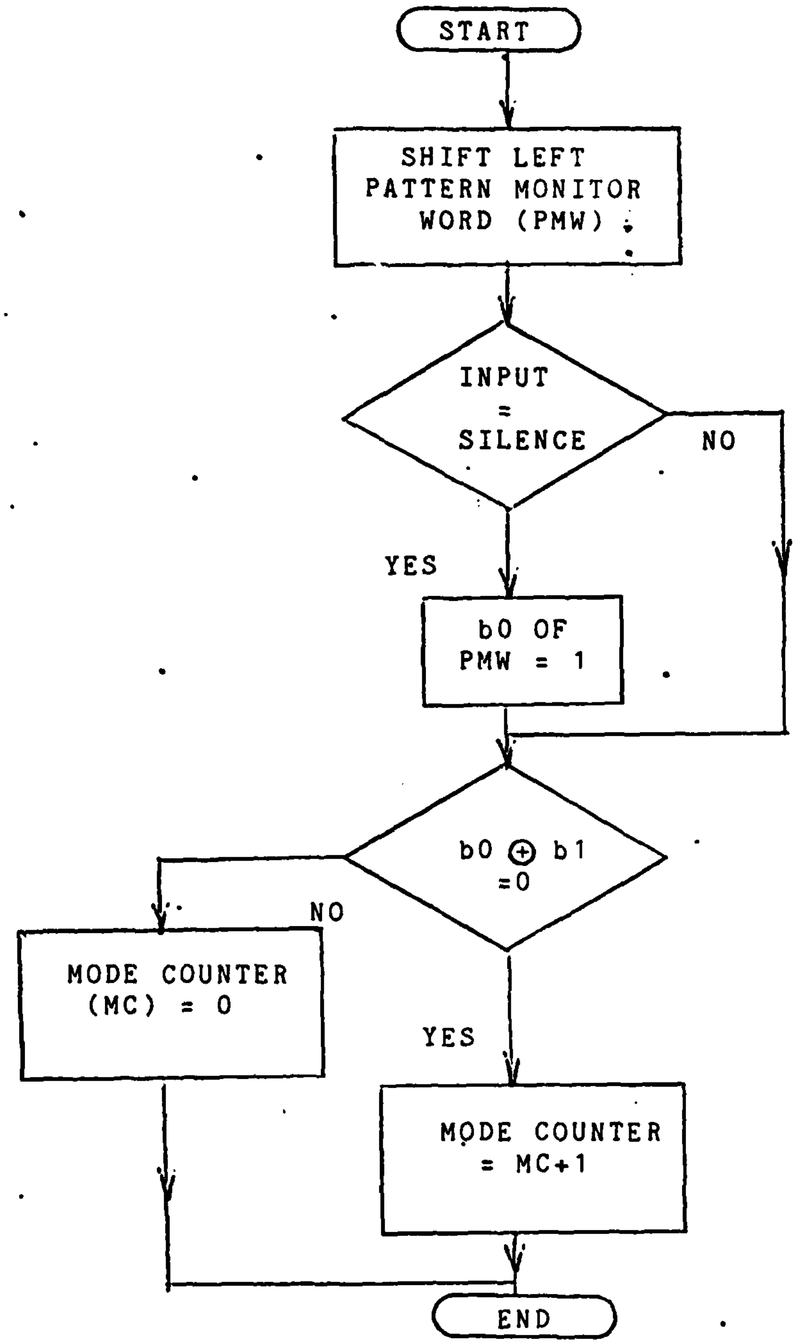


Figure 3.4.2.1 Flow diagram of Pattern Monitor word (PMW) s/s scheme

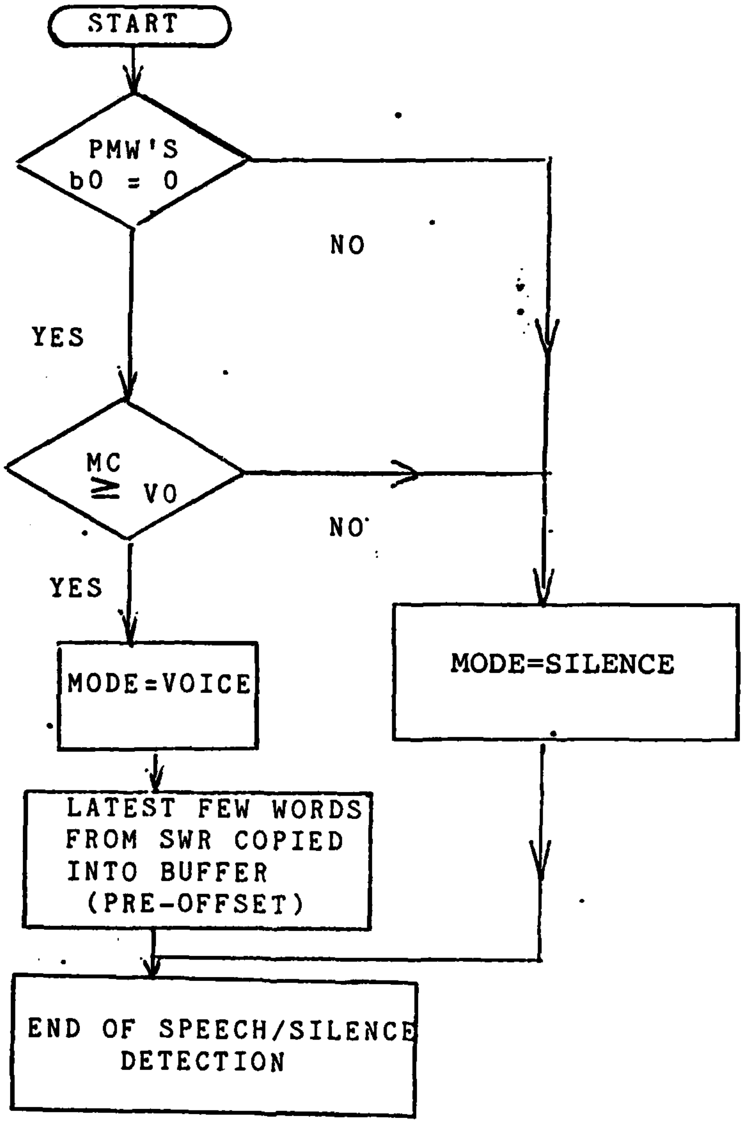
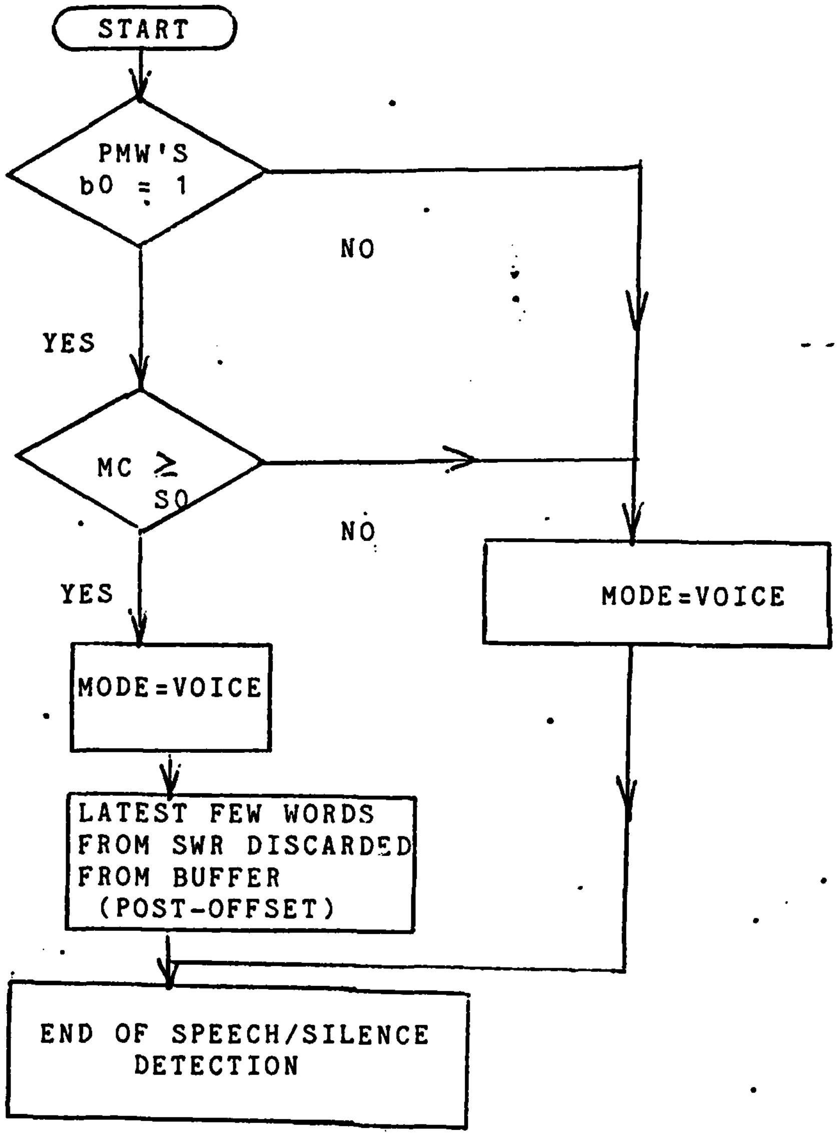


Fig. 3.4.2.2 Speech detection for PMW scheme



Figu. ..4.2.3 Silence detection for PMW scheme

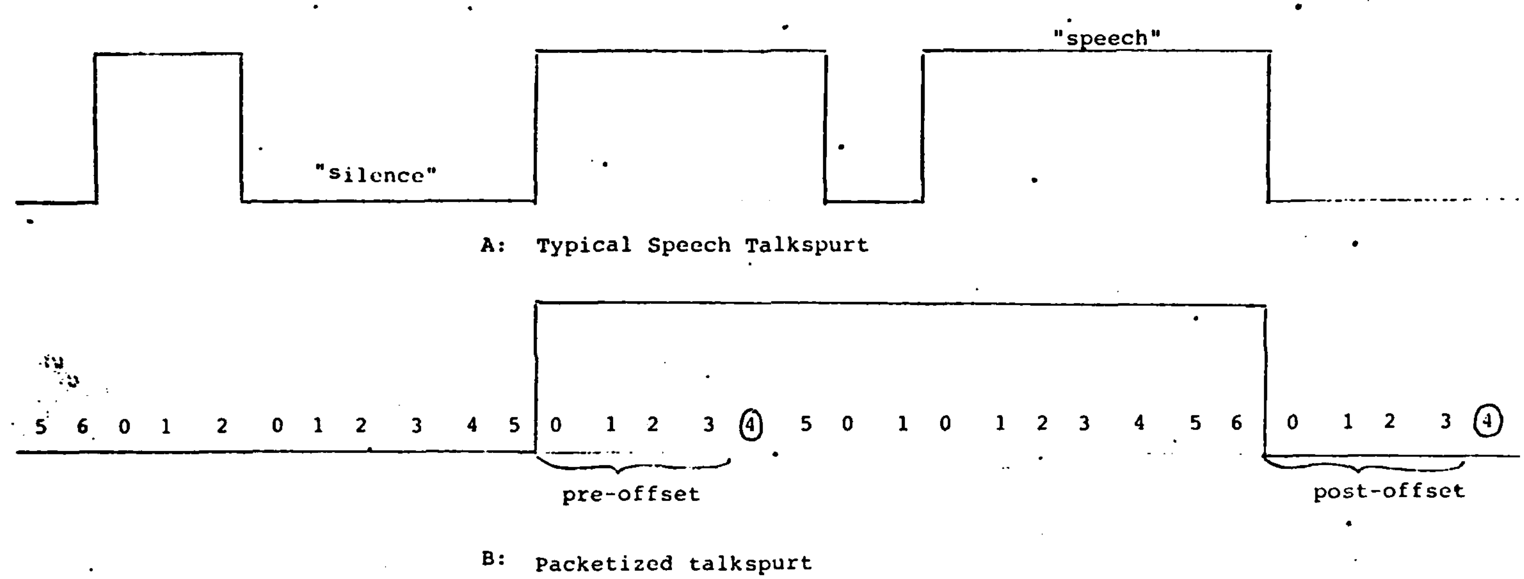


Figure 3.4.2.4 Packetization of PMW scheme of a talkspurt

Chapter 4

Packet Voice Network Simulator

Results

In this section experimental results of the packet voice network simulator (PVNS) are presented.

As discussed in chapter 3, the silence/speech (s/s) detection schemes are extremely important for efficient utilization of the communications subnetwork's facilities. To examine the efficiency of the two s/s detection schemes discussed in chapter 3, only one channel of the PVNS was utilized. Two recordings were made from radio transmissions of an all talk format station. The first recording was of a male host interviewing two male guests. The second recording was of a male host conversing over a telephone with various male speakers.

To get an accurate measurement of the packetization process, the hardware delta modulator that was used (system used is as shown in figure 2.3.1) was modified. The usual minimum step size was changed from 10mv to 40mv. This was due to the fact that the recordings made from the radio transmission were noisy and to have a 4 volt p-p signal the silent periods exhibited a 35mv noise margin. Since the

10mv minimum step size would interpret the 35mv noise signal (silence) as speech, the minimum step size was therefore changed, to avoid constant packetization of the input waveform.

Figures 4.1 thru 4.8 show statistical results obtained for both the packet size and packet rate transmission of both s/s algorithms, as discussed previously, by the PVNS.

Figures 4.1 a and b show the percentage of the transmitted packets size for the 16 bit s/s detection algorithm for various values of S_0 and V_0 for the radio male/male radio interview with the noise margin (silence) reduced to 20mv. Figures 4.2 a and b show the corresponding transmitted packet rate which is a percentage of the maximum packet rate for the various thresholds that were used. For example a maximum packet size of 1024 bits, that is encoded at a 16kbps rate, would yield a packet transmission rate of 15.625 packets/sec. It should be noted that for the 16 bit s/s algorithm s_0 must be greater than V_0.

Figures 4.3 a and b and figure 4.4 are the corresponding figures to 4.1 and 4.2, however, for these the noise margin was increased to 35mv. As a result of the increase in the noise margin of the input signal the percentage of the maximum packets transmitted decreased several percentage points, while the distribution of the packet sizes 0 thru 127 bytes increased slightly. There

was an increase of approximately 5% in the equivalent packet rate transmission, as shown in figure 4.4, over the lower noise margin of figure 4.2.

Figures 4.5 a,b and c and 4.6 a and b show the same statistics as the previous figures for packet size and equivalent packet rate for the second recording (of male/males telephone conversations). These figures show similarity to the values graphed in figures 4.3 and 4.4, since the 35mv noise margin was also used in this part.

Figures 4.7 a and b and 4.8 a and b show the transmitted packet size and equivalent packet rate statistics for the word-by-word s/s detection algorithm using 35mv noise margin. These show a marked increase in the maximum packet size transmitted of as much as 8 to 10%. There is also an increase in the equivalent packet rate of approximately 5% for this algorithm. The thresholds used for this s/s algorithm, unlike the 16 bit s/s algorithm, can vary relative to each other, e.g. $S_{\emptyset}$ can be smaller than $V_{\emptyset}$, since the counter (PMW) used is reset to zero after a change in the input is discerned (from speech to silence pattern and vice versa).

To show the usefulness of the PVNS in determining the optimal thresholds, vis-a-vis the packet size and packet rate, curves of the quality of the transmitted packetized speech is used. The graph of quality of the packetization process versus the thresholds used is shown in figures 4.9

and 4.10: Using figure 4.9 as an example [16 bit algorithm was used for the m/m radio interview (and 35mv noise margin)]. The quality was rated from poor to excellent as the various thresholds were changed and with the pre-offset, post-offset and packet loss all equal to zero.

For the 16 bit s/s detection algorithm, the optimal threshold parameters for $V_{\emptyset}$ and $S_{\emptyset}$ were chosen to be 8 and 30 respectively. These values were established using figures 4.4 and 4.9, for minimum packet rate and optimal quality. The pre-offset was set equal to 8 bytes (equal to 4msec of the initial speech) and the postoffset equal to 0.

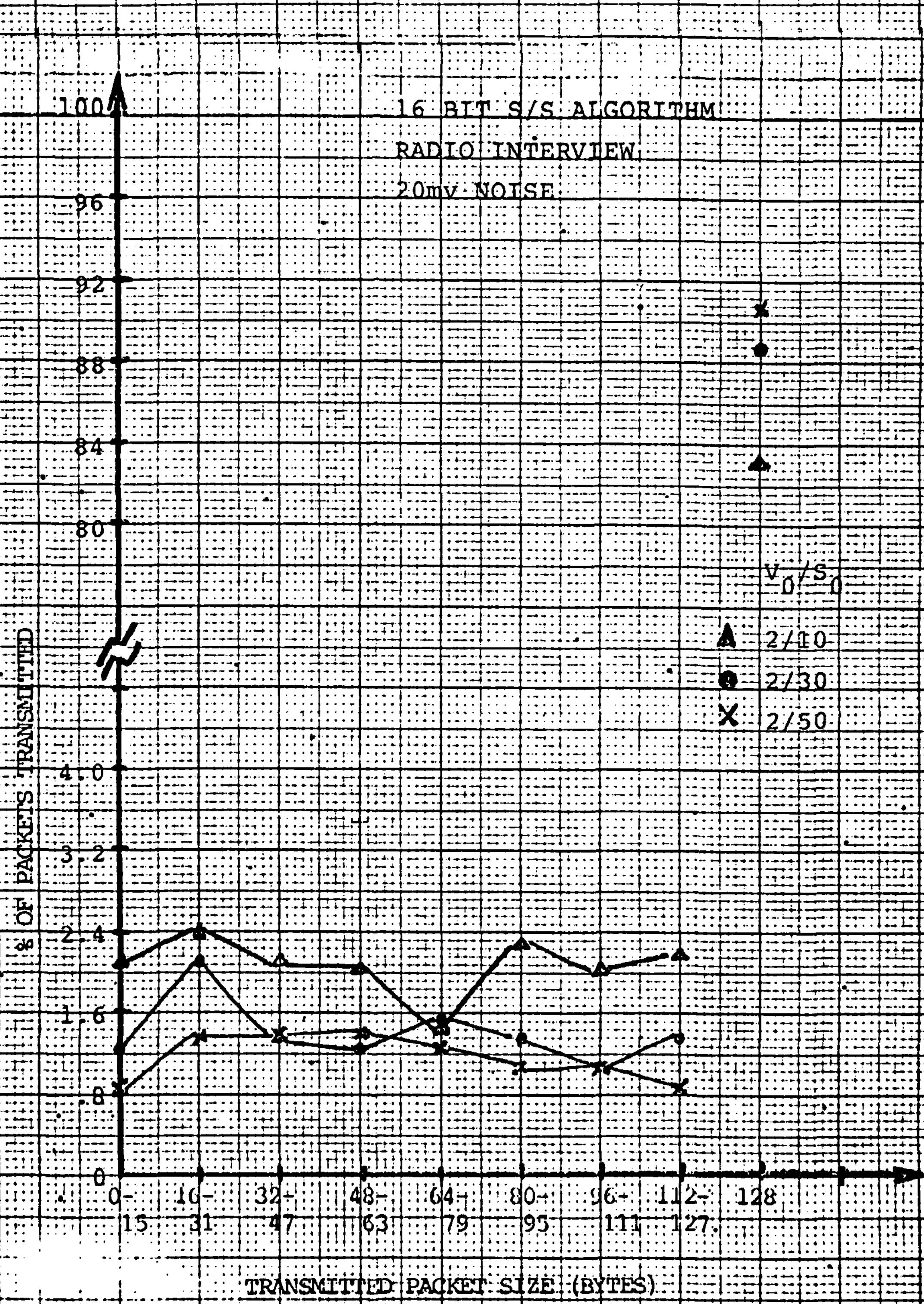


FIGURE 4.1a Transmitted packet statistics for radio interview (20mv noise)

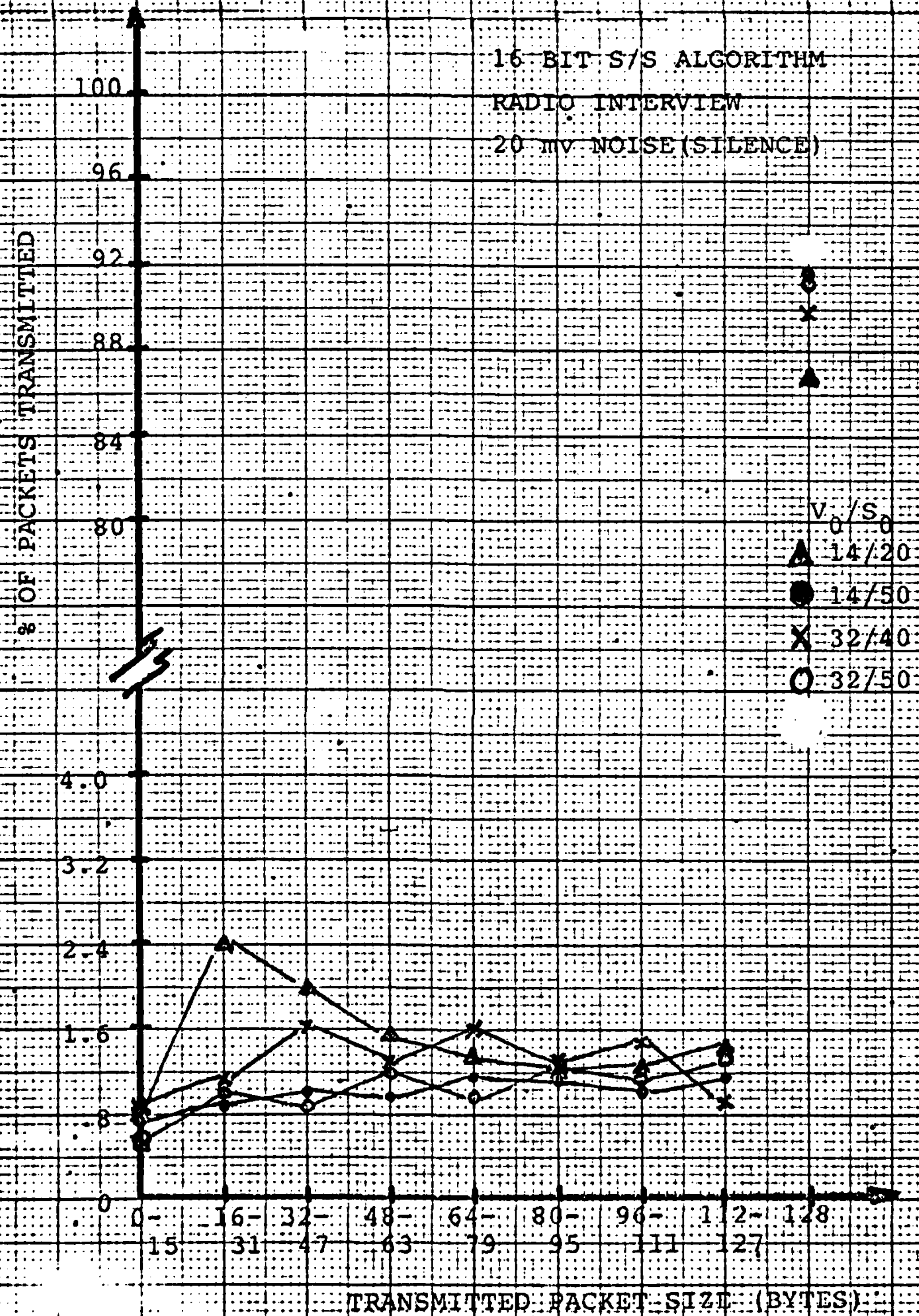


FIGURE 4.1b Transmitted packet statistics for for radio interview(20mv noise)

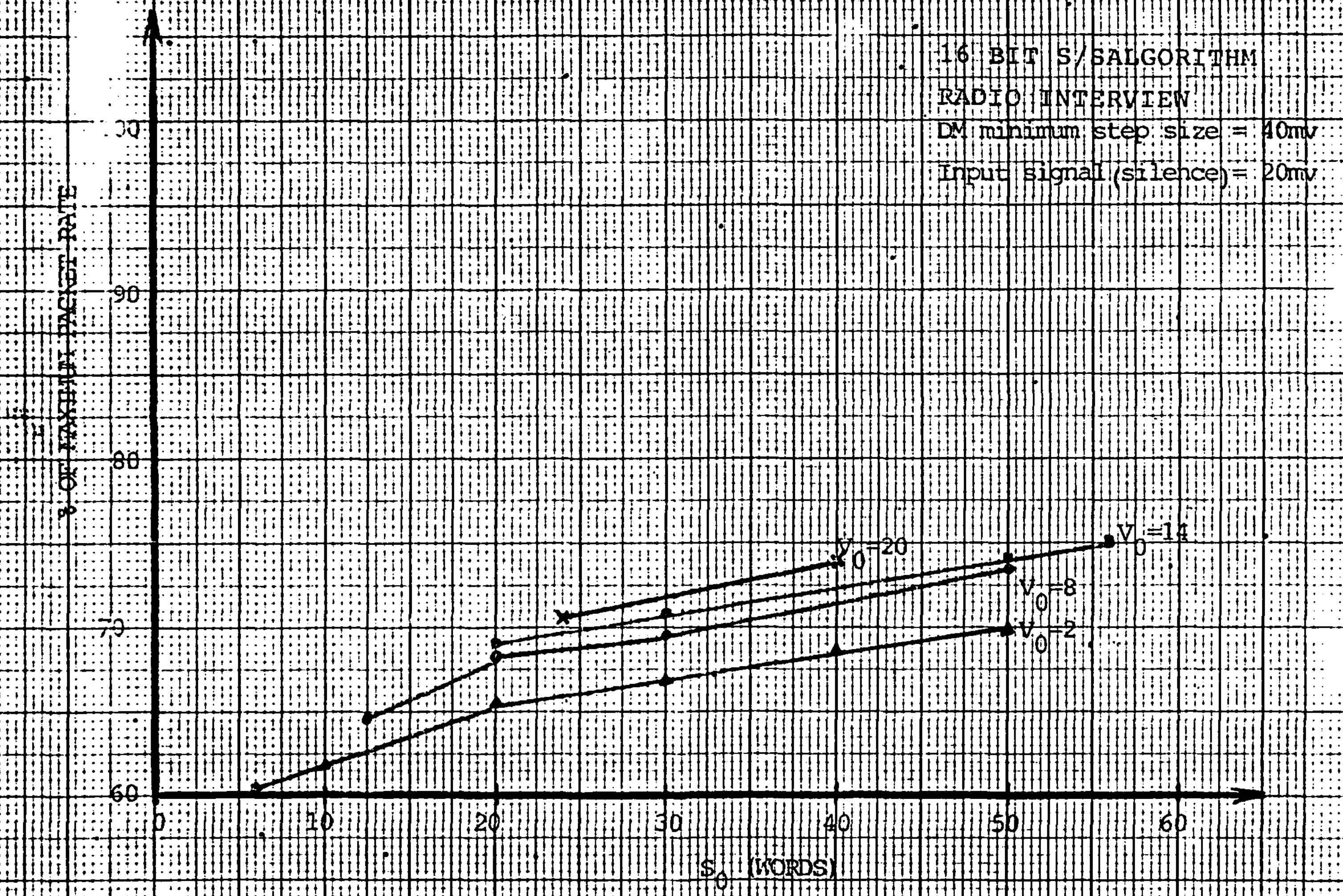


FIGURE 4.2a Packet rate statistics for radio interview (20mv noise)

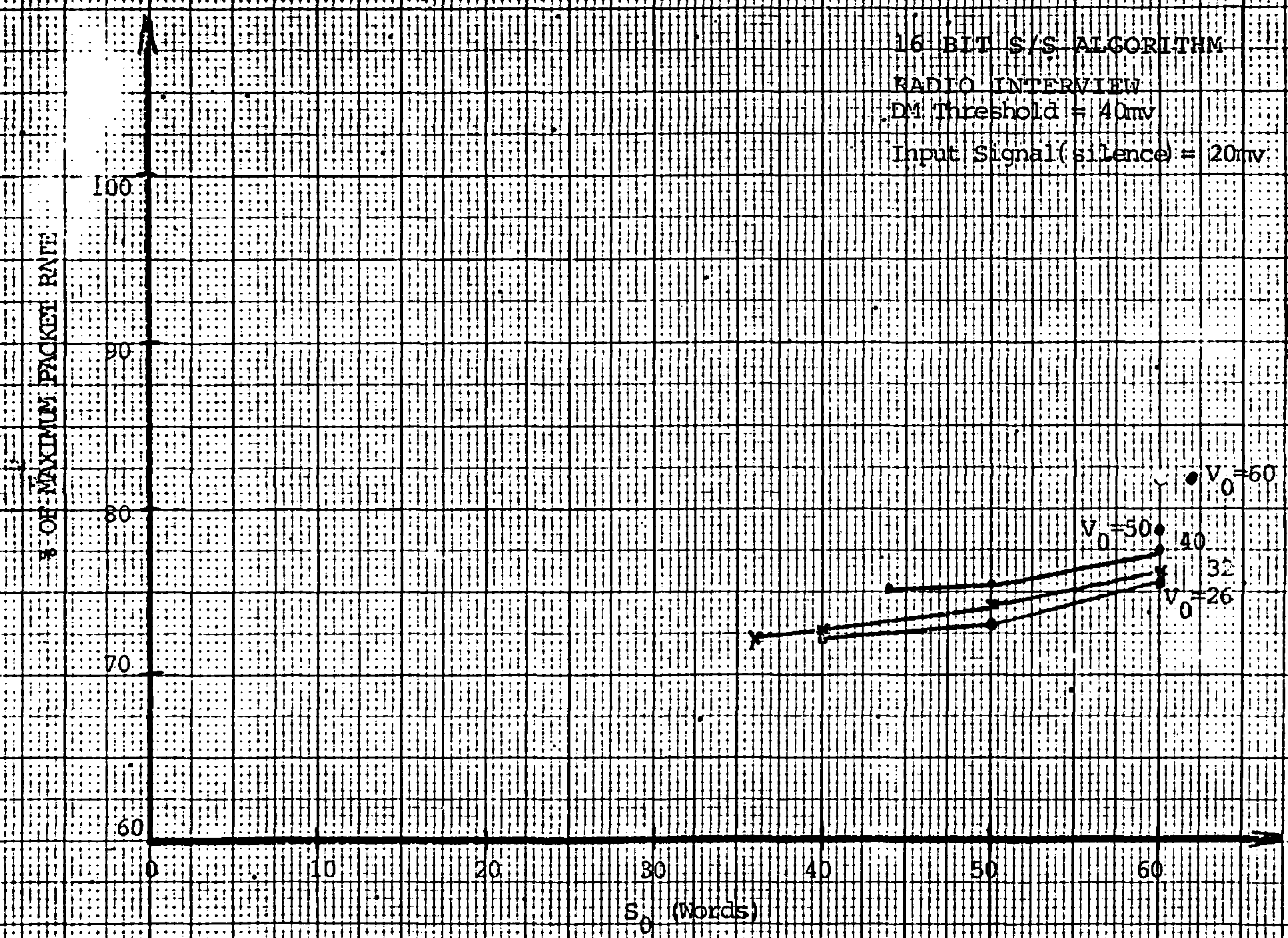


FIGURE 4.2b Packet rate statistics for radio interview
(20mv noise)

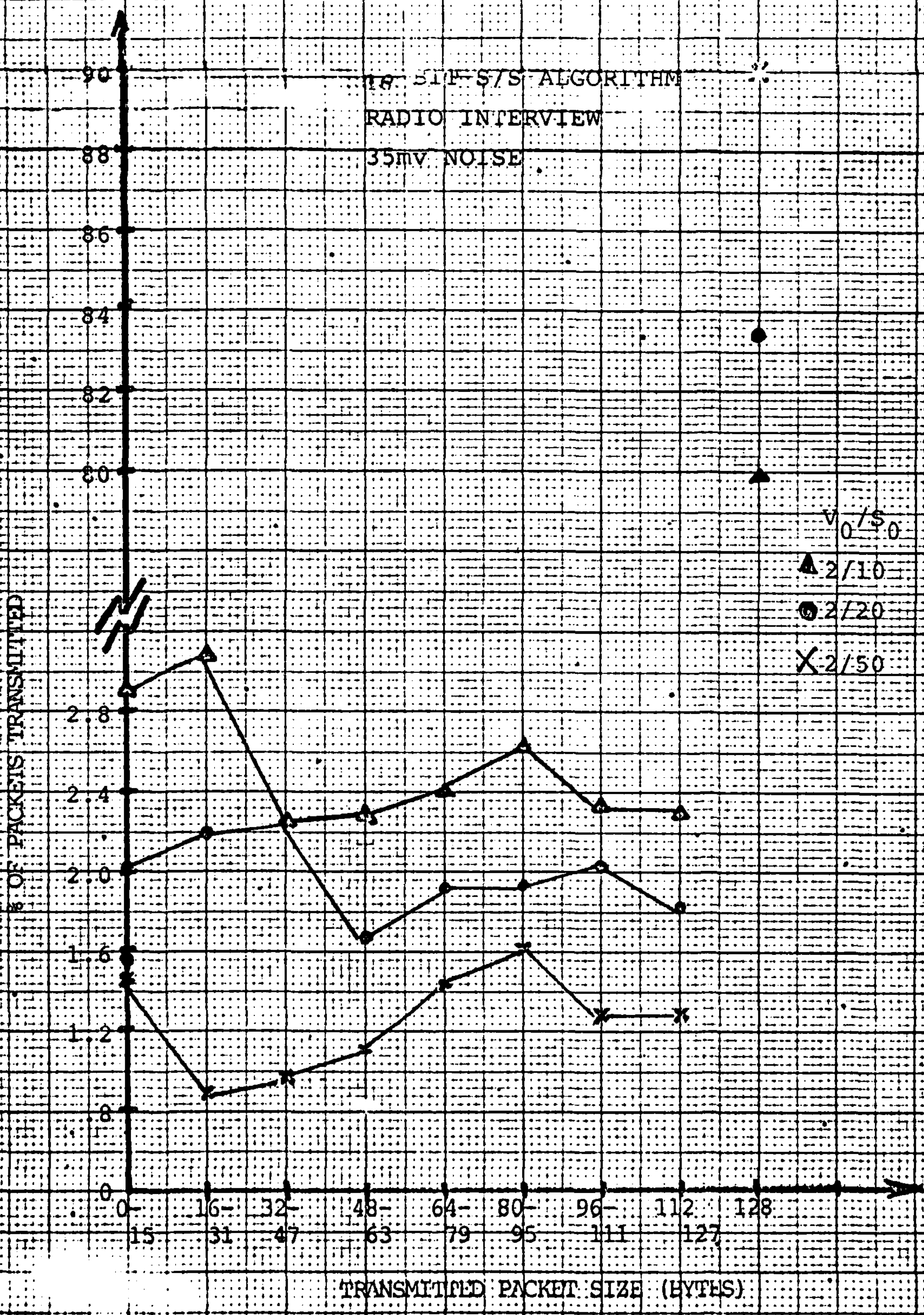


FIGURE 4.3a Transmitted packet statistics for radio interview (35mv noise)

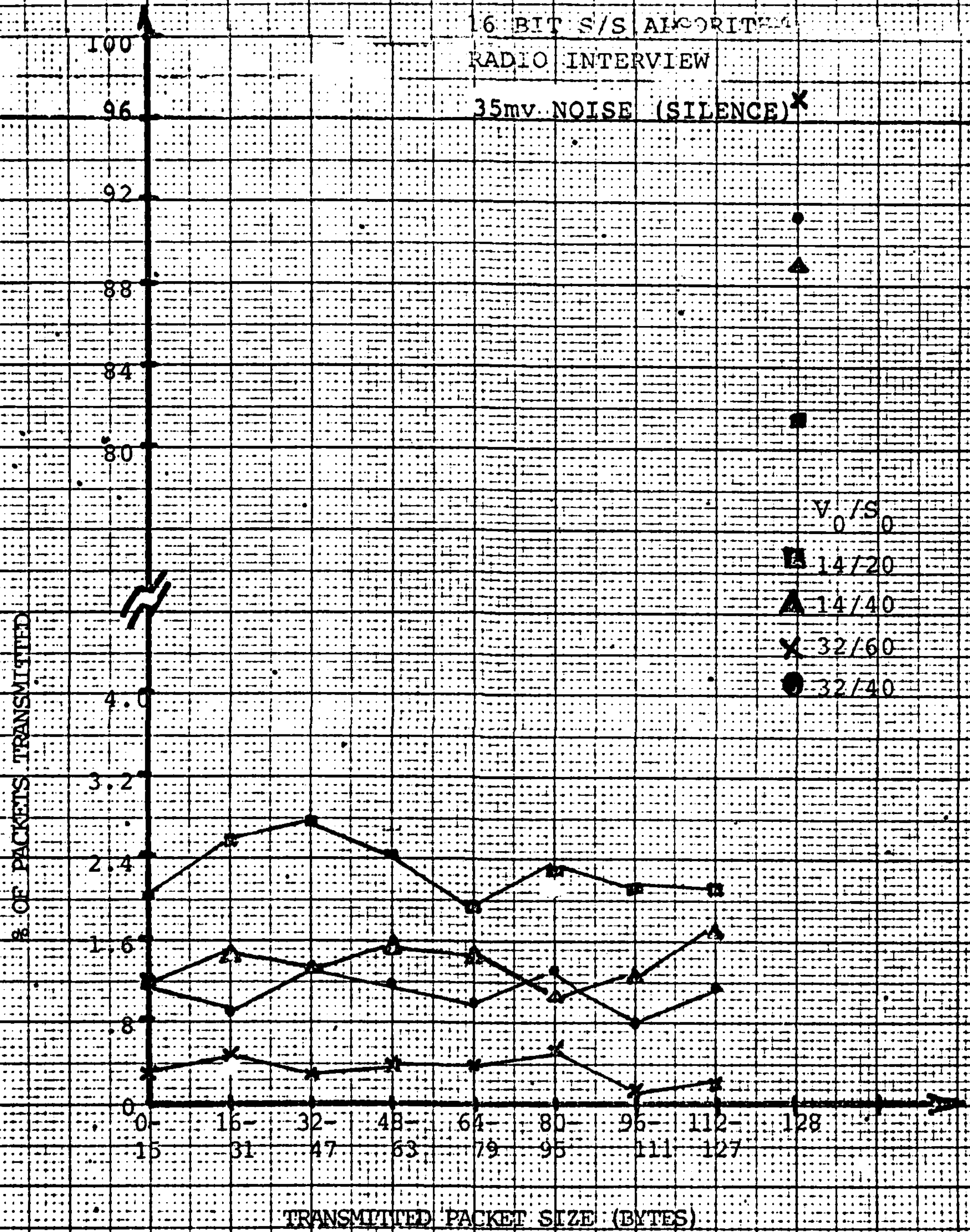


FIGURE 4.3b. Transmitted packet statistics for radio interview (35mv noise)

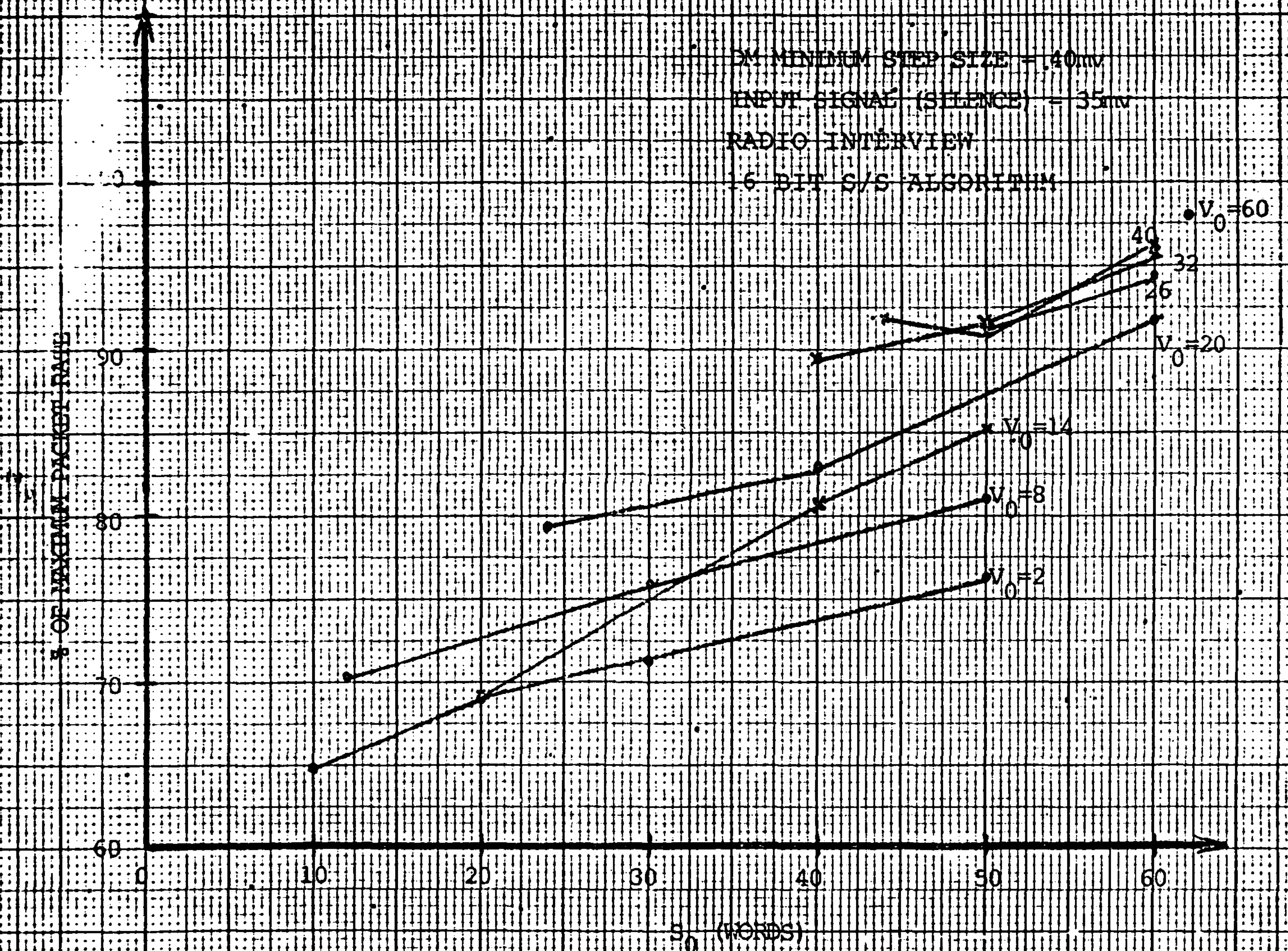


FIGURE 4.4 Packet rate statistics for radio interview (35mv noise)

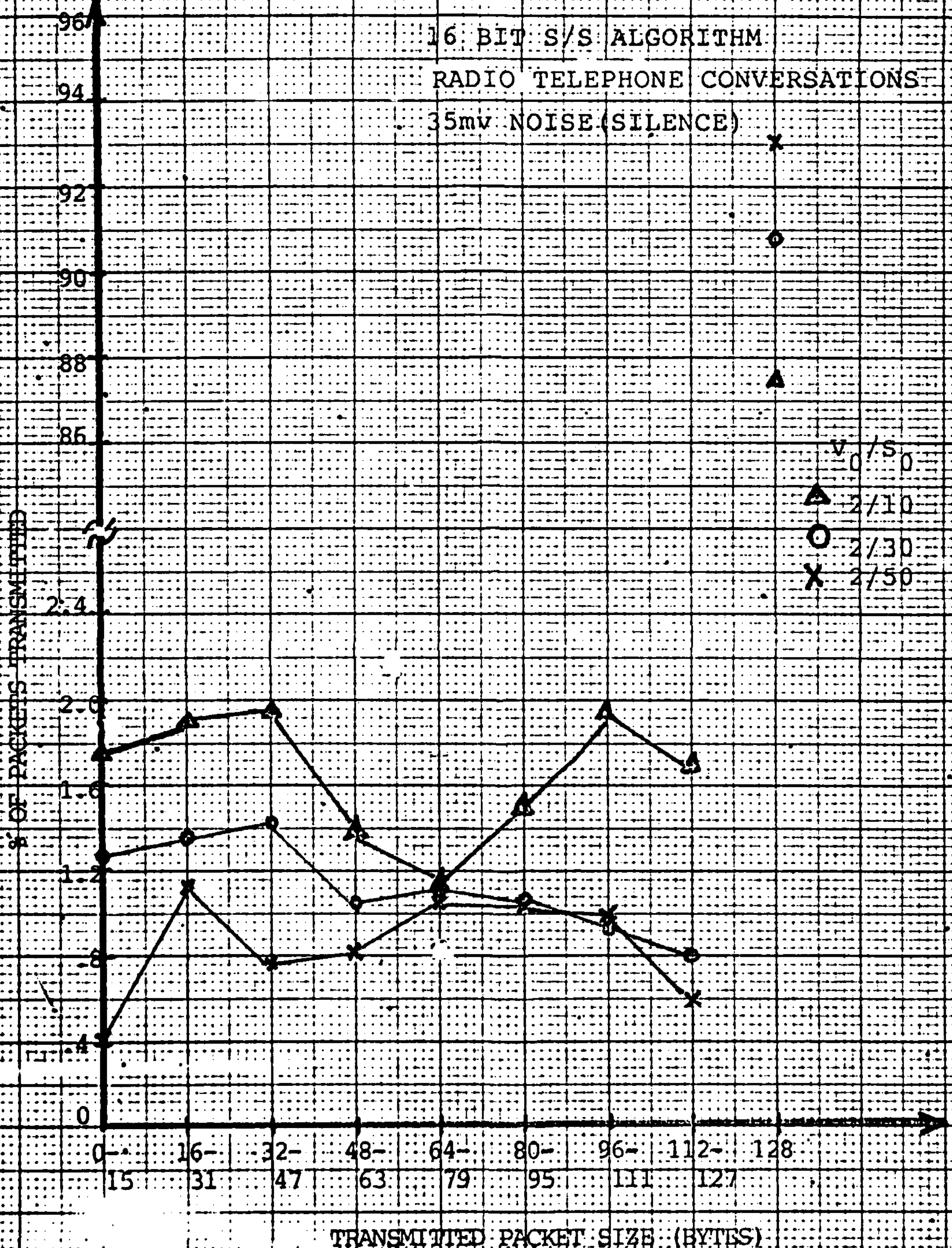


FIGURE 4.5a Transmitted packet statistics for radio telephone conversations (35mv noise)

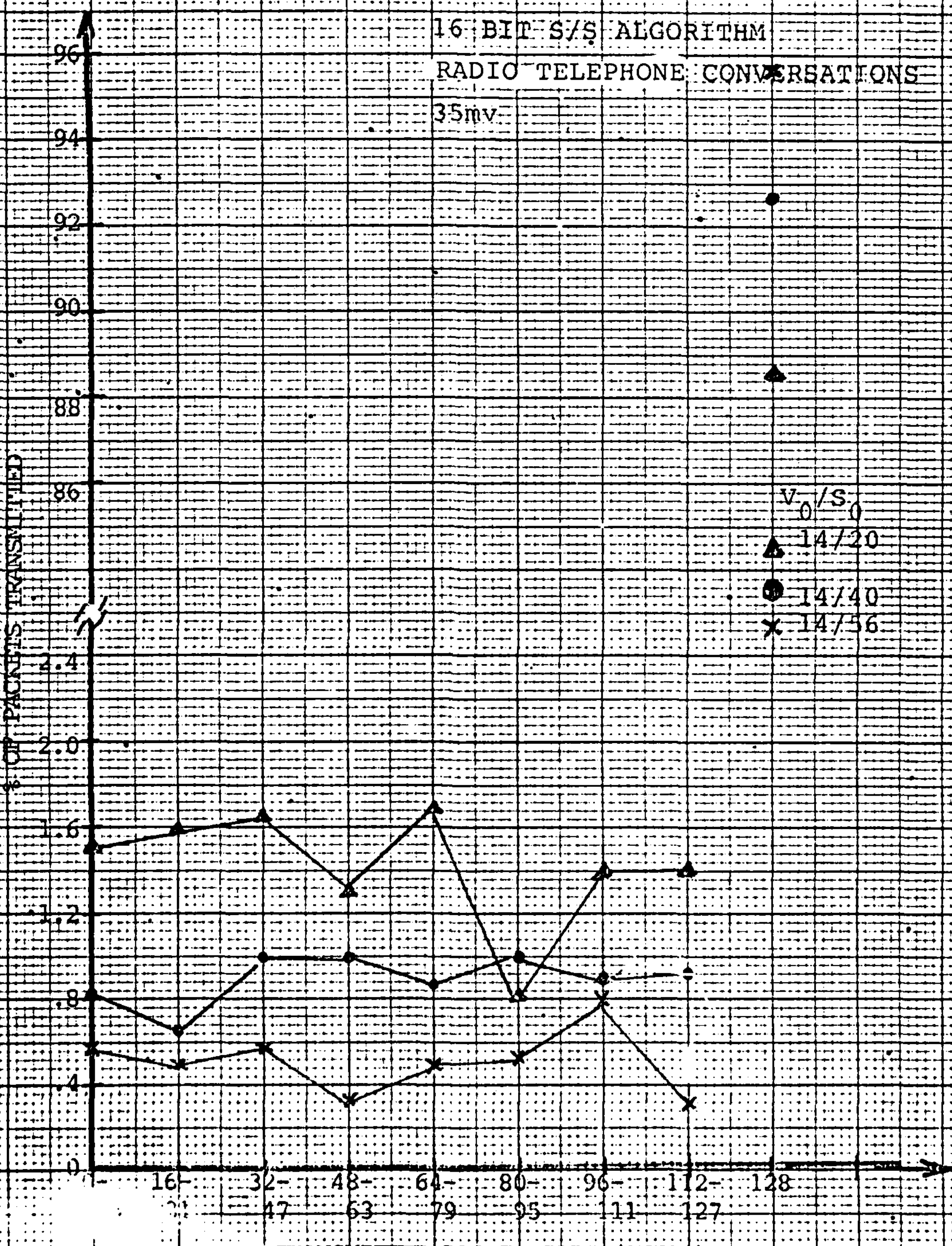


FIGURE 4.56 Transmitted packet statistics for radio telephone conversations (35mv noise)

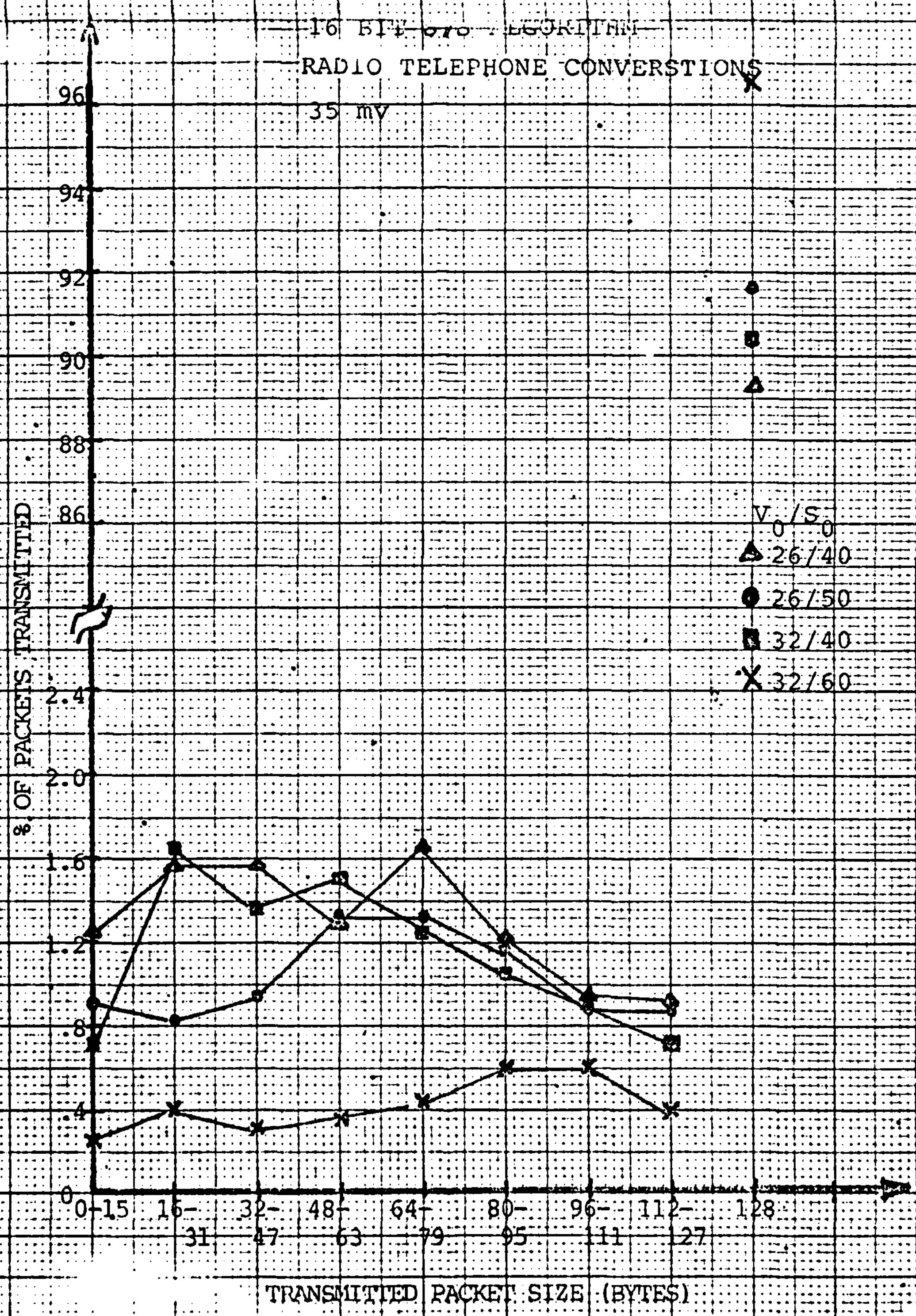
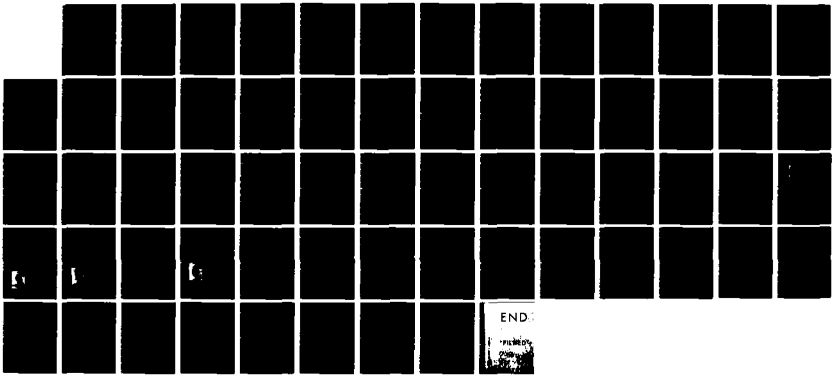


Figure 4.5c Transmitted packet statistics for radio telephone conversations (35mv noise)

AD-A133 145 LOW RATE TRANSMISSION OF VIDEO SIGNALS USING ADAPTIVE DELTA MODULATION(U) CITY UNIV OF NEW YORK RESEARCH FOUNDATION J BARBA ET AL. 15 AUG 83 MDA903-80-C-0476 2/2

UNCLASSIFIED F/G 17/2 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

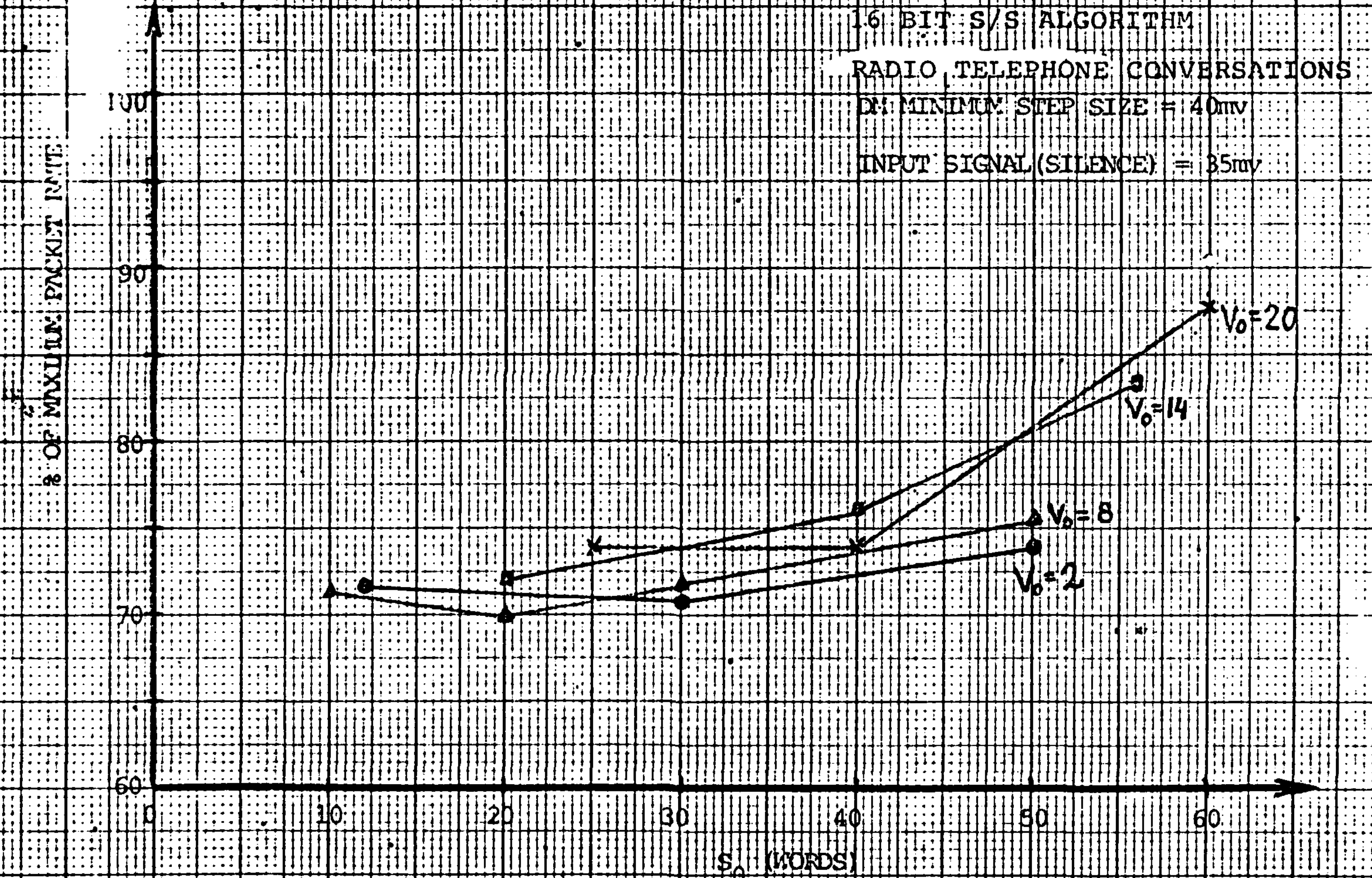


FIGURE 4.6a Packet rate statistics for radio telephone conversations (35mv noise)

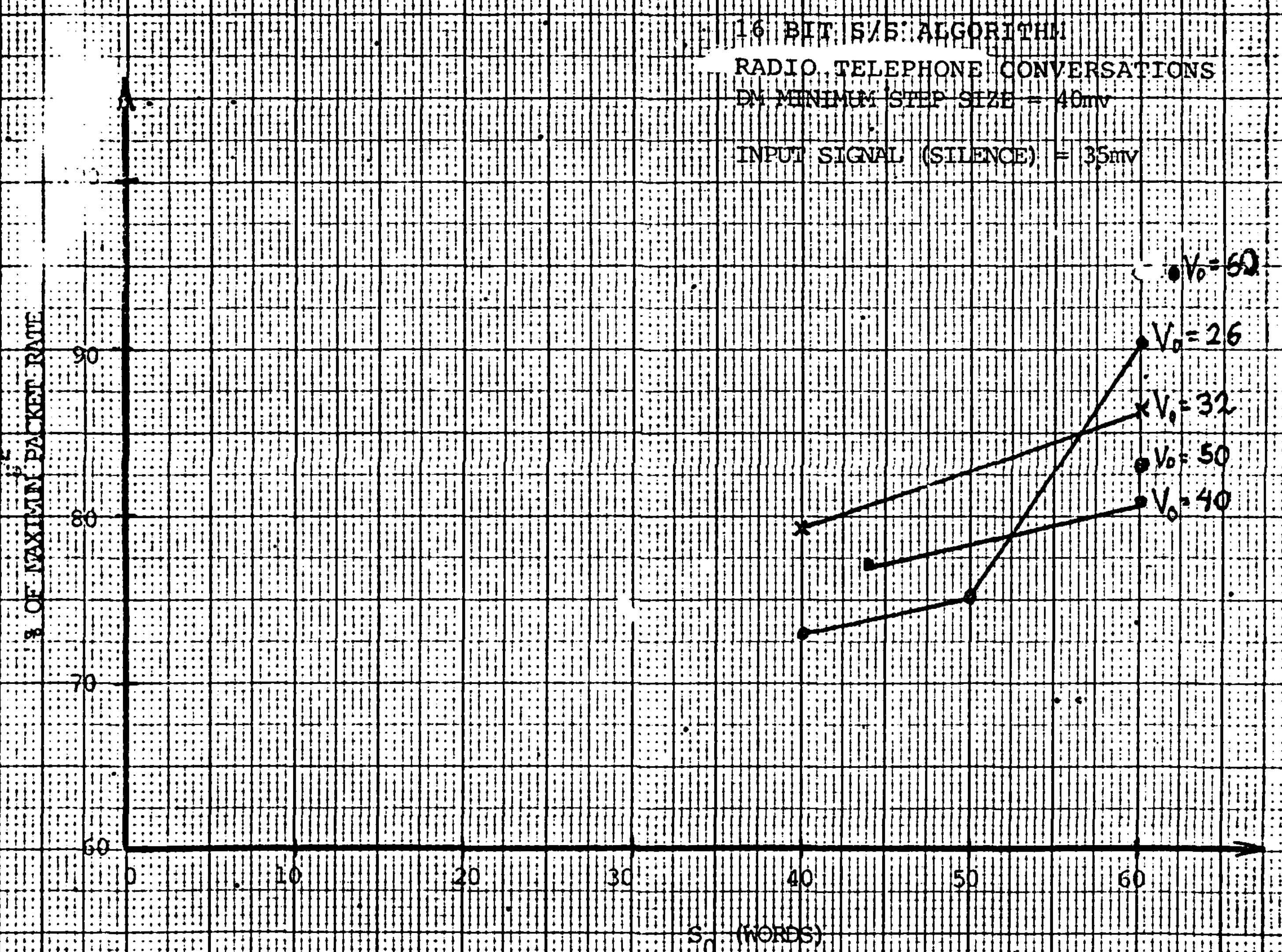


FIGURE 4.6b Packet rate statistics for radio telephone conversations (35mV noise)

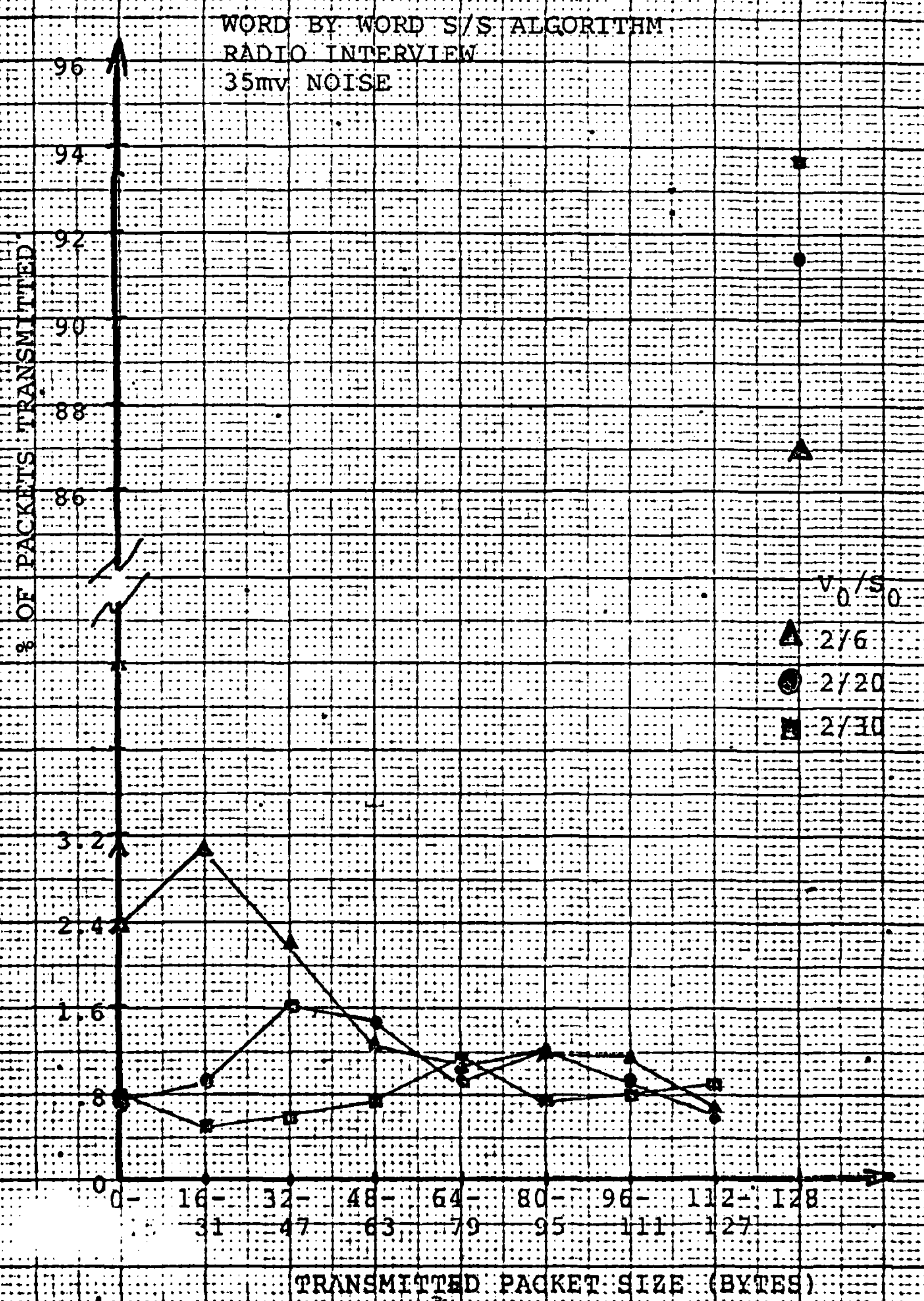


FIGURE 4.7a Transmitted packet statistics for radio interview (35mv noise) word-by-word algorithm.

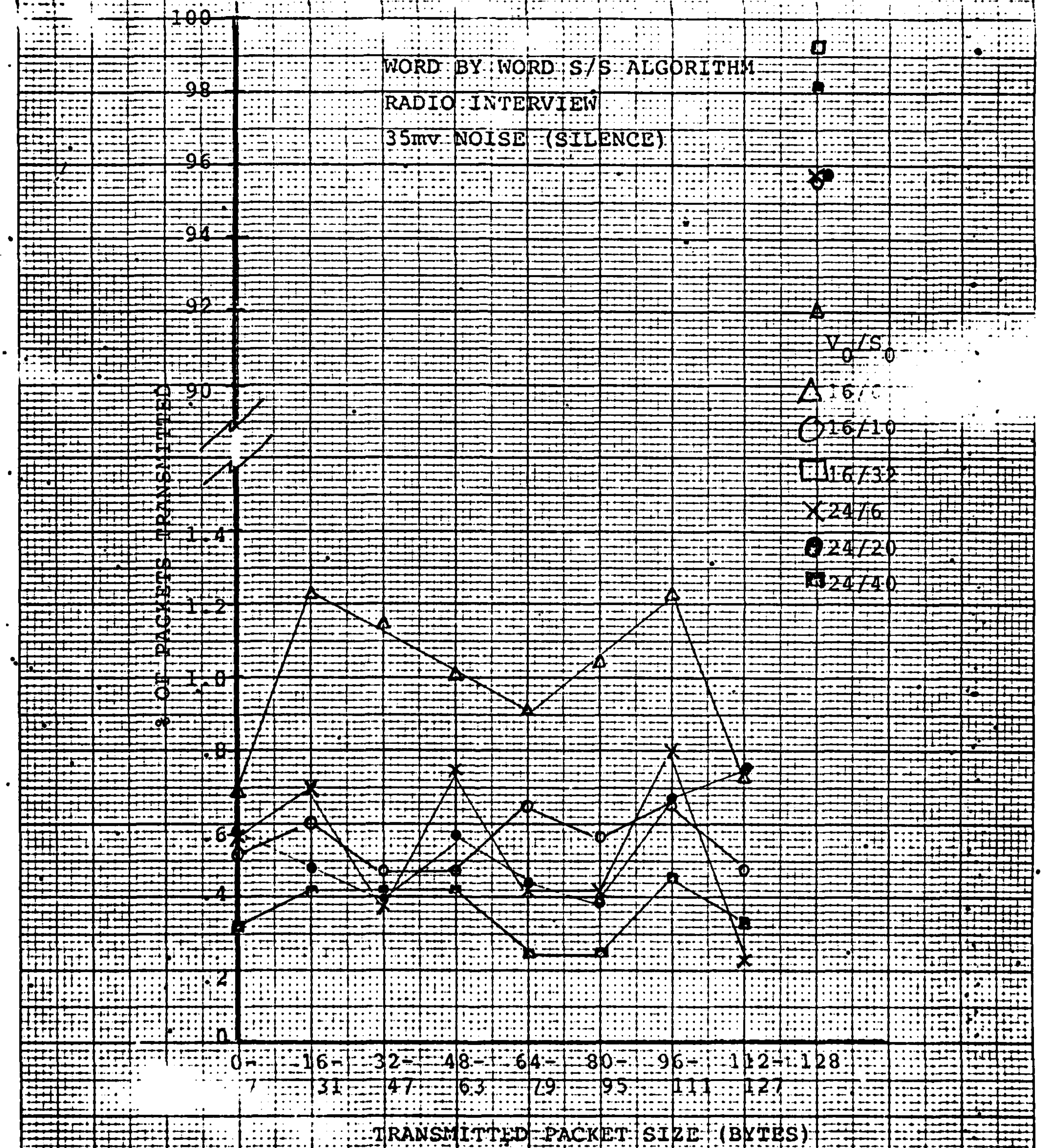


Figure 4.7b Transmitted packet statistics for radio interview (35mv noise) word-by-word algorithm

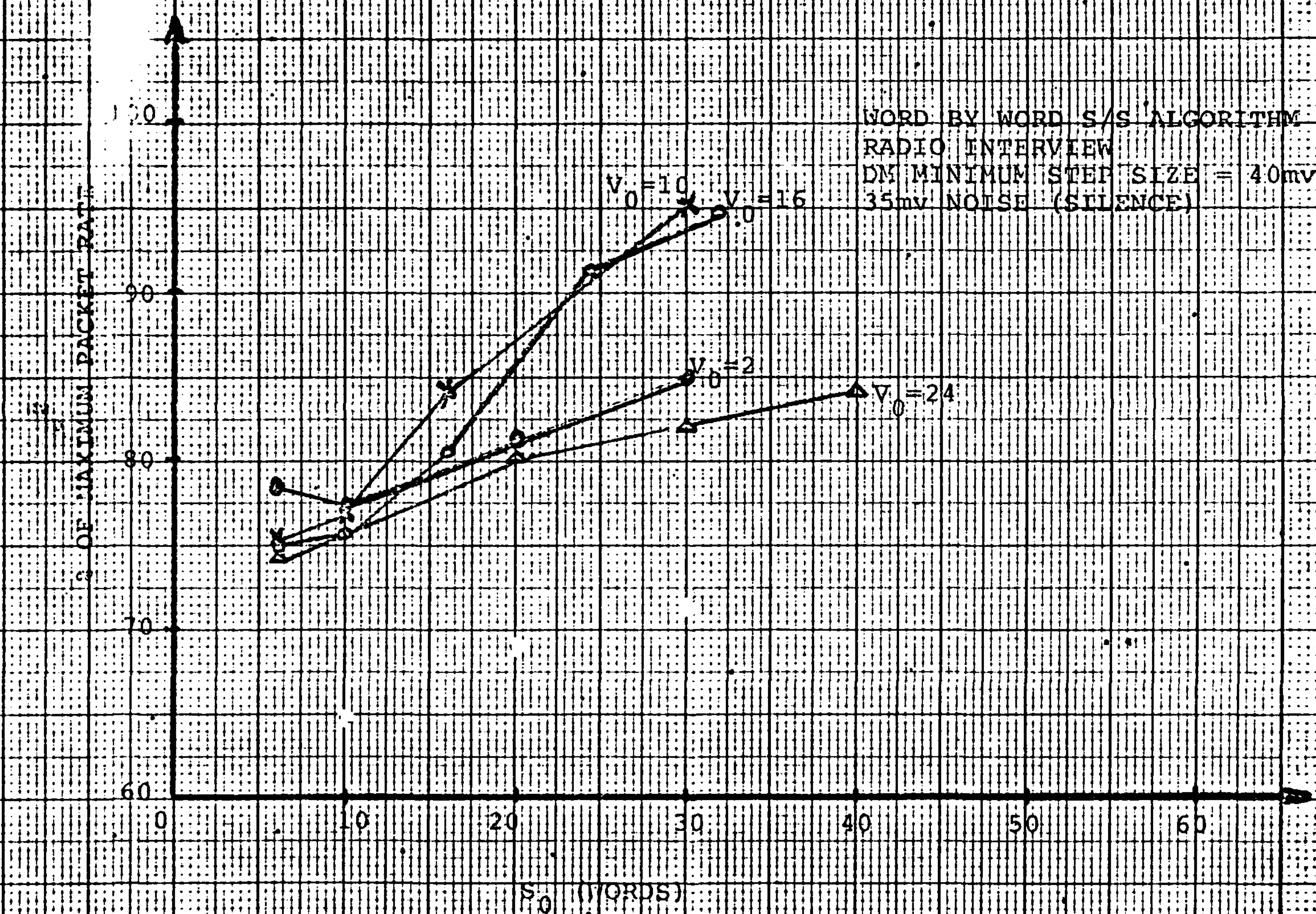


FIGURE 4.8a Packet rate statistics for radio interview (35mv noise) word-by-word algorithm

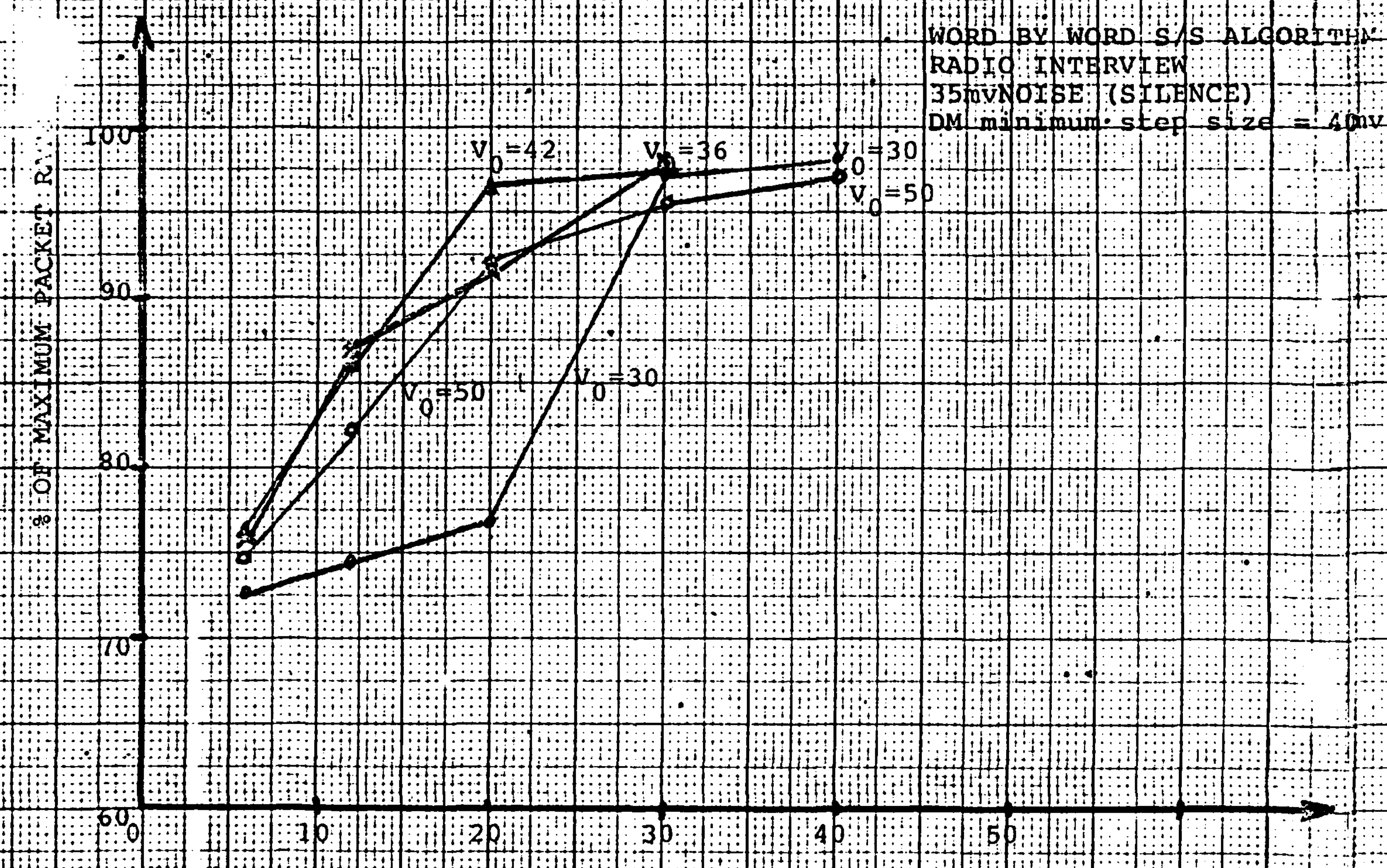


FIGURE 4.8b Packet rate statistics for radio interview (35mv noise) word-by-word algorithm

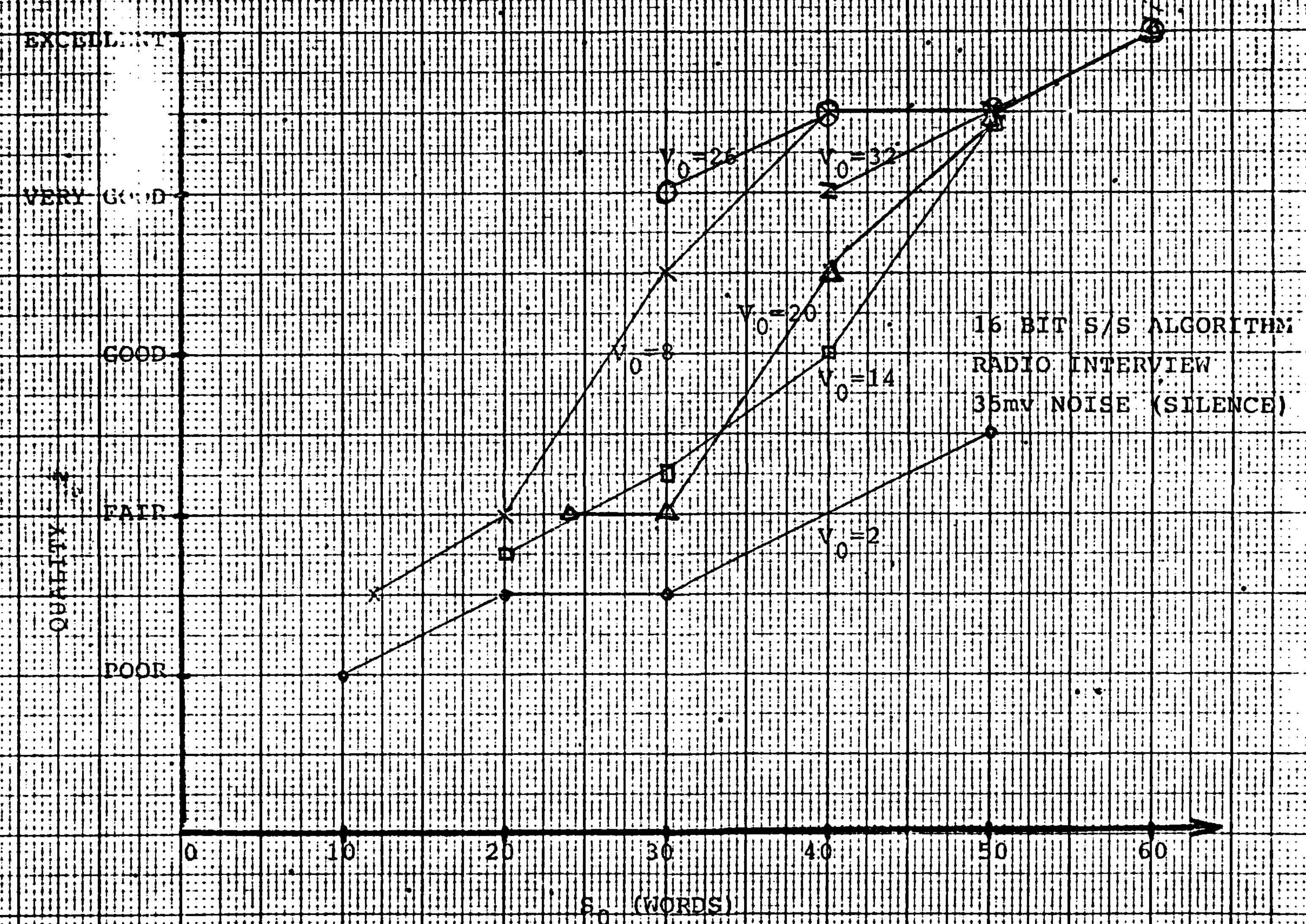


FIGURE 4.9 Quality of packetization process using 16 bit s/s algorithm

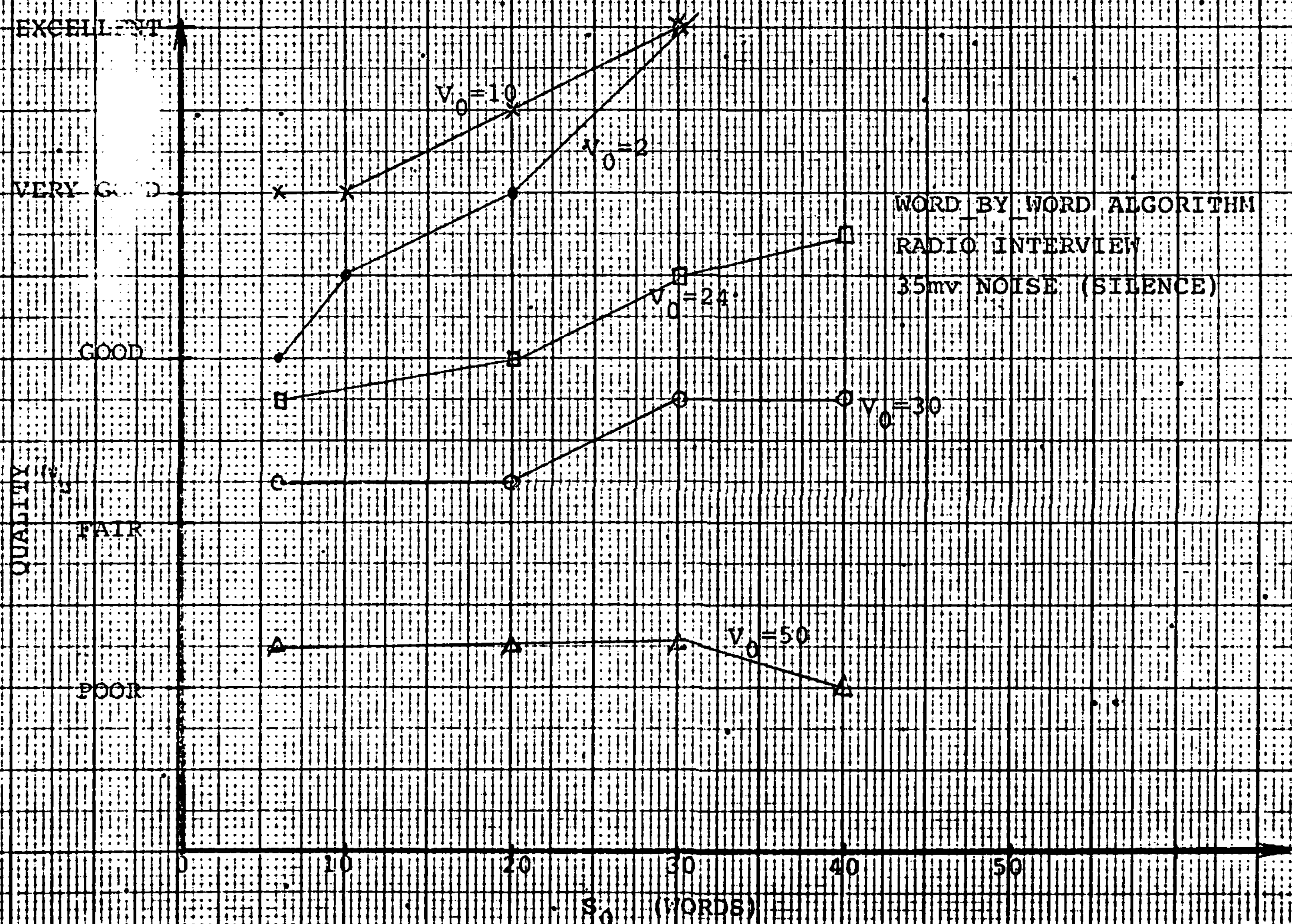


FIGURE 4.10 Quality of packetization process using word-by-word algorithm

Chapter 5

VOICEBAND M EM/SPEECH WAVEFORM IDENTIFICATION USING DELTA MODULATION

In this section we examine the performance of adaptive delta modulation (ADM) encoding for identification of voiceband data signals and speech. The application of the ADM codec in voiceband data waveforms/speech identification will be discusssed for use in the automatic routing of signals over telephone nets, with the conservation of bandwidth as the goal.

Signal identification of voiceband data or speech entails the use of the digital output of the ADM encoder to determine the autocorrelation (or equivantly the spectral characteristics) of voiceband modem signals. It has been demonstrated by Jayant [18] that when a DM is used to encode speech, the spectrum of the digital output of the DM approximated that of the input speech waveform. It will be shown that this approximation also applies when the input is a voiceband modem signal.

The autocorrelation of the ADM'S digital output and the respective spectral characteristics of various modems (2400 bps 4-phase, a 4800 bps 8-phase, a 9600 bps 16 point QAM and [illegible] bps duobinary) are determined experimentally. Measurement time considerations are addressed and

probability of error in signal identification will be discussed.

For this study the Song Voice Adaptive DM (SVADM) was used as the encoding algorithm. From here on DM will denote the SVADM algorithm.

5.1 SYNCHRONOUS VOICEBAND MODEMS

To transmit data over voiceband channels it is necessary to MOdulate the input data at the transmitter and DEModulate the received analog waveform. Thus the term MODEM evolved to describe a data transceiver. The modem in addition to translating data between data terminating equipment also performs various control functions to coordinate the flow of data between the transmitting and receiving ends.

The class of modems is divided into two general groups. These are the asynchronous and the synchronous modems. The asynchronous modems are operated at lower speeds and are used over switched telephone lines. The synchronous modems operate to speeds of 56000 bps and are used over private lines.

Synchronous modems of importance to this study and that [illegible] used for the transmission of data, operate at speeds greater than 1[illegible] bps and less than 9600 bps. These voiceband modems are employed over private voice grade

lines (usually band limited to approximately 4Khz) and use various forms of digital modulation techniques to transmit the input data stream. Table 5.1.1 lists various modems that are currently available along with the modulation techniques that are primarily used.

Voiceband modems are complicated devices that employ spectral filters, pulse shaping filters and channel equalization to achieve system requirements for proper operation. Asynchronous modems, which use start/stop codes operate below 1800 bps rate and use Frequency Shift Keying (FSK) type modulation. FSK is the preferred modulation scheme in these cases because of its simplicity of operation and ease of implementation. In FSK bandwidth efficiency is not an important factor and the bandwidth (in Hz.) is usually twice the maximum bit rate of the input data stream. The transmission of data by asynchronous modems is usually of one character (a collection of a few bits) at a time and is used interactively, such as in a time sharing environment.

Synchronous modems usually use Differential Phase Shift Keying (DPSK) modulation and combined phase amplitude schemes such as 16 point QAM. Figure 5.1.1 shows various constellations that can be used to encode the input data stream. These constellations include 4-PSK (offset QPSK), 8-PSK, two forms of 8-QAM and 16-QAM. DPSK is a modulation technique that is commonly used for the 4 and

8-PSK type modems. Figure 5.1.2 shows how a data stream (for a 2400 bps modem) converts, by some established rule, the phase of the analog waveform from its present state by a dibit pair that is presented to the modem by the data terminating equipment.

DPSK as a form of modulation requires moderate bandwidth. The term differential implies that the next symbol that is to be transmitted depends on the change in phase from the previous symbol. This change in phase is not with respect to an initially established absolute phase reference and allows for less restrictions on the synchronization at the receiver to the transmitted phase, for the demodulation of the signal.

AM is presently being used to transmit multilevel symbols by vestigial or single sideband modulation. A class of these modems transmit 4 or 8 level VSB signals and are limited to bit rates less than 10000 bps by channel impairments and power constraints. Quadrature AM, which combines both phase and amplitude for the signal set allows for greater packing of bits per symbol. The use of QAM is of importance when the signalling set uses a large number of symbols and when the average signal power is to be minimized for a minimum seperation of the states. For example, 16-QAM will have a better Probability of Error performance than 16-PSK, and is therefore the preferred encoding technique for 9600 bps modems. An alternative to

transmitting single side band is used by QAM. QAM is transmitted using two double sideband signals in quadrature. Since double sideband signals have no quadrature components, there is no interference between the two channels.

A method to transmit data at intermediate speeds, between those obtained by two and four-level systems, involves transmission of three-level signals. This type of signalling is referred to as duobinary [25]. Duobinary, for example, uses two three-level symbols to provide nine states ($3^2=9$). Thus it is possible to encode 3 bits of data into eight of the nine states provided. In multilevel systems signals can take on M values. M=2 corresponds to binary and M=3 to ternary (duobinary). In binary systems M is a power of 2, e.g. $M=2^k$ implies that each symbol represents k bits of information. Higher level systems achieve data rate packing of k bits/sec/Hz, or k times the data rate capability of binary (a 2400 bps duobinary rate (over a 3-level system) is equivalent to a 3600 bps binary transmission, therefore, providing 3/2 (1.5) times greater packing of bits/symbol).

5.1.1 TYPICAL MODEMS

The 2400 bps synchronous modem uses 2 bits of data from the input data stream at one time and according to an

arbitrary decision making process changes the phase of the carrier, as shown in figure 5.1.2. Since 2 bits of data are taken at one time, the dibit pairs allow 4 possible choices (e.g. 2^2 points or the following dibit pairs 00, 01, 10, 11) for the phase to change by from the present phase of the sinusoid. This encoding results in a transmission rate equal to 1200 bps and is referred to as the baud rate.

BIT RATE = BAUD RATE * NO. OF BITS/SYMBOL

In the 4800 bps and higher speed synchronous modems it is also possible to use a combination of amplitude and phase modulation or just PSK as shown in figure 5.1.1. The 4800 bps modem uses 8 different points to phase encode the carrier.. This results from using 3 bits of data to encode (e.g. 2^3 points or the following tribit combinations 000, 001, ..., 110, 111) the phase. In this case the transmission rate is reduced by 1/3 (symbols/bit) to 1600 baud.

The typical 9600 bps modems use combined amplitude and phase modulation technique known as 16 point QAM. These 16 points result from combining 4 bits at one time (e.g. 2^4 points or the following 4-bit combinations 0000, 0001, ..., 1110, 1111). The combination of 4 bits to transmit one phase change of the carrier result in a baud rate of 2400

bps.

Other modems operating at 3600 bps and 7200 bps are available, although these are generally not used. There are also modems operating to 56 Kbps, these are known as wideband or group modems. The wideband analog modems require a wider bandwidth than that available on a voice grade telephone line, and are not of interest in this study.

5.1.2 SIGNAL FILTERING AND SHAPING

Filtering is used in modems simply to confine the signal to a specific frequency band, inorder to minimize the influence of noise. Along with filtering signal shaping is also used to help control intersymbol interference.

Although the long haul telephone path is filtered, local paths to and from the local switching centers have a wide and unrestricted bandwidth allowing crosstalk to occur over a wide frequency range. This crosstalk is detrimental to the transmission of data and must be overcome.

Nyquist described a type of spectral shaping to avoid intersymbol interference. Figure 5.1.2.1 shows a baseband signal having a rectangular spectral shape limited to W Hz. This rectangular pulse will have a corresponding (sinx)/x time function with zero crossings at T=1/2W time

intervals. Thus it is possible to transmit pulses at a 2W rate without the peaks of adjacent pulses interfering with each other. Since it is impossible to to generate functions with infinite cutoff, Nyquist and others developed other practical forms of spectral shaping that would have the same zero crossing points. The modifications to the square pulse is shown in figure 5.1.2.2. These spectral shapes have odd symmetry about the cutoff frequency as indicated in the figure. The additional bandwidth required to transmit the signals with these spectral shapes is expressed in terms of the originally discussed bandwidth W. The time response that corresponds to 100% rolloff (full $\cos^2$ rolloff) has a property that in addition to the T=n/2W crossover points (where n=1,2,...) there is no interference at the half amplitude points. The time response also dies away more rapidly than the 0% and the 50% rolloff slopes.

Modems used in telephony are designed with a raised cosine shaping of the spectral density (SD). The SD can be described mathematically at baseband as follows [22].

$$S(w) = \begin{cases} T, & 0 \leq w < (\pi/T)(1-\alpha) \\ T/2*[1-\sin(T/2\alpha)*(w-\pi/T)], & (\pi/T)(1-\alpha) \leq w \leq (\pi/T)(1+\alpha) \end{cases} \tag{1}$$

Where T in the above formula represents time. $\alpha=1$ corresponds to a full raised cosine spectrum while $\alpha<1$ results in a spectrum with a flat portion at low frequencies and raised cosine shaping at the edges. The multiplication of the above discussed baseband spectrum, by a carrier, shifts the waveform to the frequency of the carrrier. Typical spectral characteristics of modems are shown in figure 5.1.2.3. It should be noted that spectral shaping of the voiceband data signal can be done at baseband before modulation or after modulation and can also be accomplished as a combination of the two.

5.2 EXPERIMENTAL PROCEDURE

The experimental procedure used to calculate the Autocorrelation (or spectral density) characteristics of various modems is as follows:

The data applied to the digital input of the modem was a binary PN sequence generated by an HP 3722-A noise generator. The noise generator was clocked by a signal provided by the modem, which is at the bit rate of the moder [illegible] is used. The analog voiceband modem signal is sampled by the DM encoder which is clocked by an independent clock source. The digital output of the DM

encoder was then input into a PDP 11/34 mini-computer by a DR11-K parallel interface bit-by-bit and is shown in figure 5.2.1.

The autocorrelation (R) of the digital output was then calculated using ensemble averaging techniques. The autocorellation was averaged over (N=) 65,535 independent measurement intervals. Each measurement interval included (n=) 100 bits. The autocorrelation was calculated as follows:

$$R(m) = \frac{1}{N}\sum_{m=1}^{N} e_k(0)e_k(m) \qquad m=1,2,...,n \qquad (2)$$

Where $e_k(j)$ is the jth bit of the kth measurement interval.

The Spectral Density (SD) was calculated using the following formula derived by Bennett [19].

$$W(f) = \frac{1}{T}|G(f)|^2 * \{R(0)-m_1{}^2+2*\sum_{k=1}^{n}[R(k)-m_1]^2\cos(2\pi kfT)\}$$
(3)

Where m_1 is the mean and $R(0)$ is the correlation of every bit with itself. In this study this formula above

redu^^ to:

$$W(f) = \frac{1}{T}|G(f)|^2 * (1-2\sum_{k=1}^{n} R(k)\cos(2\pi k f T)) \qquad (4)$$

Where f=1/T is the DM sampling rate, and G(f) is the Fourier transform of a unit pulse of width T.

$$G(f) = \sin(\pi f T)/\pi f T) \qquad (5)$$

5.3 DIGITAL TRANSMISSION VIA TELEPHONE LINES

With the present mix of Analog/Digital telephone transmission facilities, the future challenge is the use of digital transmission exclusively in data communications. Presently data and voice communications are acheived as shown in figure 5.3.1.

Figure 5.3.1a shows how voice channels are converted to digital format, for transmission over digital telephone lines. A 4Khz (=W) analog voice signal is PCM encoded as foll[illegible] The input signal is sampled at the Nyquist rate (=2W), resulting in 8Ksps (samples per second) signal. These samples are then encoded using a 7 bit A/D (=128 (2^7)

levels). To each 7 bit encoded sample an additional bit, used for timing control, is added. The resulting digital stream which is at a 64Kbps rate is time division multiplexed (TDM) along with 23 other such voice channels and transmitted over a T-1 line at a 1.544 Mbps rate. figure 5.3.1b shows that voiceband modem signals are also PCM encoded and time division multiplexed for transmission over a T-1 line in exactly the same way as described above. Figure 5.3.2 shows the format used for TDM frame of data on a T-1 line. 7 bits of data or voice per channel is multiplexed along with an additional bit for timing control with other encoded samples. The resulting frame used is 193 bits long and equals 125 msec of time per frame of digital data transmitted on the T-1 line.

The heirarchy of the telephone digital transmission system used in the U.S. is shown in fig. 5.3.3. T-1 lines at 1.544 Mbps are cross connected with other T-1 lines, also 4 T-1 lines can be multiplexed (M1-2) resulting in a T-2 line (6.312 Mbps rate). Similarly T-2 lines are multiplexed (M2-3) resulting in a T-3 line which transmits at a 44.736 Mbps rate.

5.4 EXPERIMENTAL RESULTS

The measured autocorrelation and the calculated spectra of four modems are examined, they include:

Racal/Vadic	2400 bps	4-PSK
GTE-Lenkurt	4800 bps	8-PSK
Western Electric	9600 bps	16-QAM
GTE-Lenkurt	2400 bps	DUOBINARY

The measured autocorrelation that resulted from ensemble averaging techniques, as outlined above, is shown in figures 5.4.1 and 5.4.2. Figure 5.4.1 compares the measured autocorrelation of the 2400 bps and the 4800 bps modems. The DM bit sampling rate used in these measurements was 32 Kbps. The use of this particular sampling speed was purely arbitrary, although a maximum sampling speed of only 38 Kbps was possible due to limitations of the PDP-11/34. Figure 5.4.2 shows the measured autocorrelation of the 4800 bps and 9600 bps modems also measured at a 32 Kbps sampling rate.

It is seen from figure 5.4.2, which is plotted for several independent measurements, that there is some scatter in the individual values of R(n) at all n. The scatter is small and does not affect the general shape of the curve. Figure 5.4.3 shows the calculated spectral density of the 2400 bps and the 4800 bps modems. It is obvious that the spectrum has a bandpass shaping since the autocorrelation is similar to an exponentially decaying sinusoid. The 6dB bandwidth of the plotted spectra are

approximately 1160 Hz for the 2400 bps modem and 1625 Hz for the 4800 bps mo... n. This compares with the expected values of 1200 Hz and 1600 Hz respectively.

Figure 5.4.4 and 5.4.5 show the calculated and measured power spectral density (20log amplitude) for the 4800 bps and 9600 bps modems. These curves show the calculated spectra for three different and independent measurements of autocorrrelation using n=75 bits of data over 65,535 independent measurement intervals.

Figures 5.4.6 and 5.4.7 show, respectively, the measured autocorrelation function and calculated spectra of the 2400 bps duobinary modem.

Ensemble averaging techniques used to calculate the autocorrelation function, as discussed above, necessitates long measurement time (several minutes). If the input data to the modem is assumed to be random and ergodic, then time averaging techniques should yield results equivalent to ensemble averaging techniques.

For time averaging of the autocorrelation, a long sequence of consecutive bits output from the DM is entered int the memory of the PDP-11/34 computer in the same way described previously. The autocorrelation is then calculated as follows:

$$R(m) = \frac{1}{N}\sum_{k=1}^{N} e(k)e(k+m) \qquad m=1,2,\ldots,n \qquad (6)$$

Here the sequence of input bits must have a total length equal to N_T=N+n bits.

Figure 5.4.8 and 5.4.9 show the results of the measurement of the autocorrelation using time averaging. Values of N=10,000 and N=5,000 are shown for the 4800 bps and the 9600 bps modems. The figures were plotted for five independent measurements of the voiceband data waveform. First it is seen that the general shape of these time averaged measurements is very similar to the curve obtained by ensemble averaging techniques. Further it is observed that the scatter at any specific value of n increases as N is decreased.

5.5 DISCUSSION

The observed similarity of the time averaged and ensemble averaged autocorrelation makes the previous a viable choice for real time applications. This is true because time averaged measurements take orders of magnitude less time than ensemble averaged measurements. The time involved in these measurements (at 32 Kbps) varies from seve[illegible]onds to fractions of a second. This value of the measurement time deopends on both N and the number of n's that are to be calculated.

It is the purpose of this study to examine the possibility of discriminating between various modems. To study the capability of distinguishing between various modems, using the measured (time averaged) autocorrelation function, an experiment was performed. It was desired to distinguish between the 4800 bps and the 9600 bps modems. N=10,000 bits was used and five autocorrelation values were needed. This necessitated a measurement of N_T=10,005 bits of DM digital output. The autocorrelation bits that were used were the 16th-19th and 23rd-24th. The value of each R(n) was added in the following way:

$$Th=\{-[n_{17}+n_{18}+n_{19}]+[n_{23}+n_{24}]-n_{16}\}$$

A value of Th equal to 2800 was used. If the threshold (Th) was greater than 2800 it was decided that a 9600 bps modem has been detected, and a value less than 2800 resulted in a decision that 4800 bps modem is on line.

The experiment was performed 10^5 times and the results are shown in table 5.5.1. A probability of error less than 10^{-5} was desired and values of 4.4 10^{-4} and 2.7 10^{-4} resulted from the measurements performed on the modems. These results are encouraging because if the voiceband modem waveforms are sampled at speeds of 64 Kbps and great[illegible] smaller values of N can be used, therefore less time is needed to complete the measurements. It is also

true because the faster the sampling rate the better the DM can follow the waveform. Figure 5.5.1 shows the autocorrelation of the 9600 bps modem sampled at 25 Kbps and 35 Kbps. It is seen that there is just a shrinking in the waveform, this observation again implies that at higher sampling rates (64 Kbps and greater) would result in similar curves and would achieve error rates less than 10^{-5}.

An experiment to distinguish the 4800 bps and 9600 bps modems from the 2400 bps duobinary modem was also performed. For this experiment n_{26} and n_{40} were added and compared to a threshold value. Error rates of less than 10^{-5}, in distinguishing the 2400 bps duobinary from the 4800 bps and the 9600 bps modems were measured.

5.6 CHARACTERISTIC DIFFERENCES OF SPEECH AND VOICEBAND MODEM WAVEFORMS

Speech and Voiceband modem data waveforms differ radically from each other, as shown in figures 5.6.1 and 5.6.2.

Speech waveforms are an amalgm of different waveforms linked together and, as mentioned in previous chapters, speech includes long periods of silence. Voiceband modem waveforms, on the other hand are irregularly shaped sinusoidal signals with approximately equal amplitude.

Due to the long silence periods found in conversational speech, the energy associated with it is bursty and has its spectral energy concentrated below 800 Hz. The energy flow of data is generally smooth with its spectral energy spread evenly about the carrier frequency (approximately 1800 Hz).

To distinguish between speech and voiceband modem waveforms, the long silence periods of conversational speech are exploited. It was noted that the DM outputs a steady state pattern when there is silence. Therefore, to establish that there is speech on the line, the presence of the steady state pattern is monitored.

5.7 APPLICATION-AUTOMATIC ROUTING USING DELTA MODULATION

In transmitting data signals over a telephone network, it is possible for data signals to undergo at least four format conversions, as shown in figure 5.7.1. This is true because, although digital transmission facilities exist, conventional analog modems are used when a network can not be directly accessed using digital signals.

To maximize the channel utilization of present and future digital transmission facilities, an alternative using automatic routing is proposed. The ADM can be used as a tool to automatically route voiceband modem or speech

signals by examining the correlation function (or spectrum) of the ADM's digital output. The system to be used is shown in figure 5.7.2. There are various inputs to the auto-router, these inputs include both speech and voiceband modem signals. The output of the auto-router is multiplexed between a voice or data concentrator. The signals from the data and voice concentrators are then TD-multiplexed for transmission over T-1 lines. The specific elements of the proposed system are discussed below.

5.7.1 AUTOMATIC ROUTER

The schematic for the auto-router is shown in figure 5.7.1.1. The input to the auto-router can be either a speech waveform or a voiceband modem signal. The input is delta modulated by the ADM/Controller (ADM/C). The ADM/C examines the input waveform and decides on whether the input is speech or data. The controller then routs the input to a speech concentrator or to a data concentrator.

5.7.2 SPEECH CONCENTRATOR

The schematic for the speech concentrator is shown in figure 5.7.2.1. The signals that are routed to the speech concentrator by the auto-router can be handled in various ways. The speech waveform can be encoded by ADM or PCM

codecs.

If ADM codecs are to be used, the speech signal can be encoded at 16Kbps or at a 32Kbps rate, this effectively increased the channel utilization by factors of 4 and 2, respectively (PCM is encoded at 64Kbps per channel). The digitized signals are then multiplexed and routed to a TDM system for transmission over T-1 lines.

5.7.3 DATA CONCENTRATOR

The schematic of the data concentrator is shown in figure 5.7.3.1. The voiceband modem signals are routed to the data concentrator in a specific manner. Within the concentrator there are banks of specific demodulators (e.g. 2400 bps QPSK modem, 4800 bps 8-PSK modem, etc.). These banks of various demodulators decode signals from specific modems.

The digital (demodulated) signals are then multiplexed at much lower bit rates then the PCM system presently used (e.g. the modem's bit rate vs. 64Kbps PCM). The demodulated bits are then routed to a TDM system where they are multiplexed with speech signals and other modem signals for transmission over T-1 lines.

DATA RATE (bps)	MODULATION TECHNIQUES
110 - 1800	FSK
2000 - 2400	4-PSK Vestigial Sideband Duobinary
3600	4-Phase + AM
4800	4-Phase + AM Vestigial Sideband 8-PSK
7200	Phase and Amplitude Modulation
9600	Phase and Amplitude Modulation

Table - 1 Voiceband Modem Characteristics

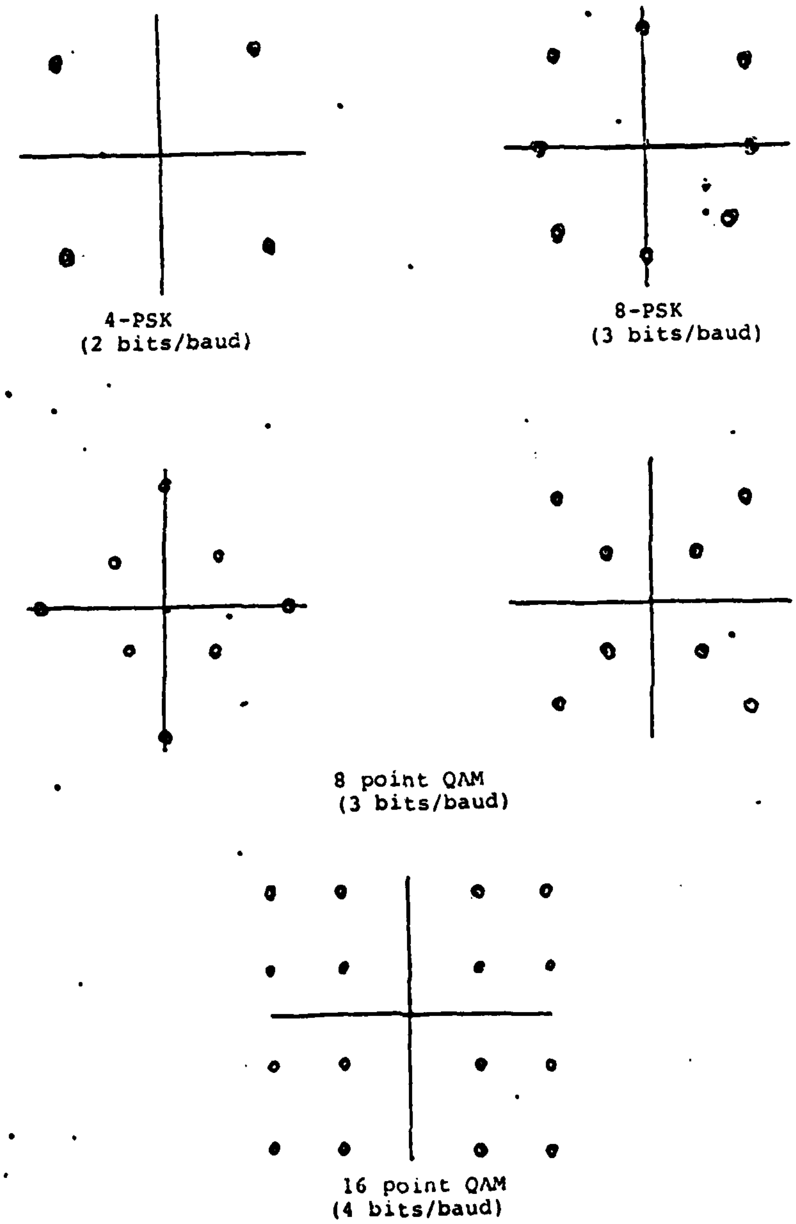
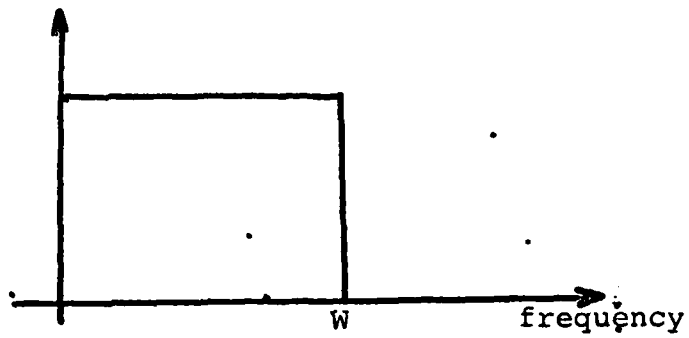
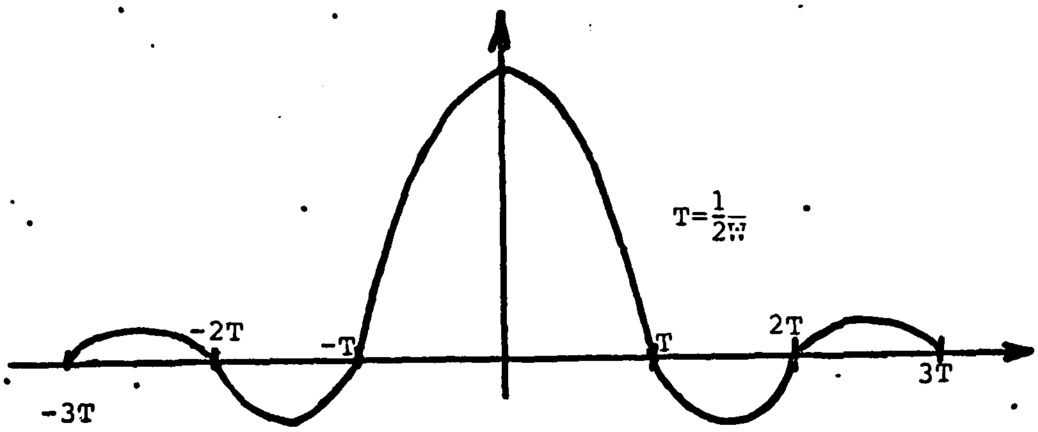


Figure 5.1.1 Various constelations used in voiceband modems



(a)



(b)

Figure 5.1.2.1 a) baseband signal with rectangular spectrum
b) time response of a

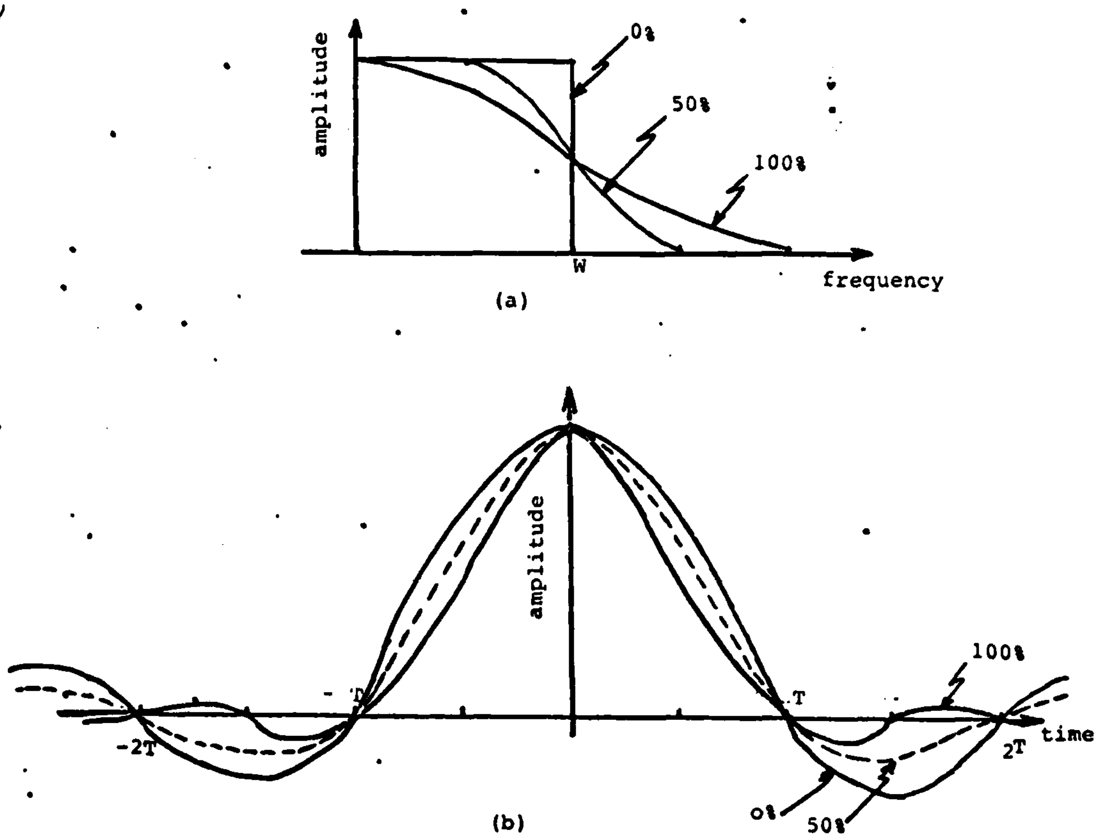
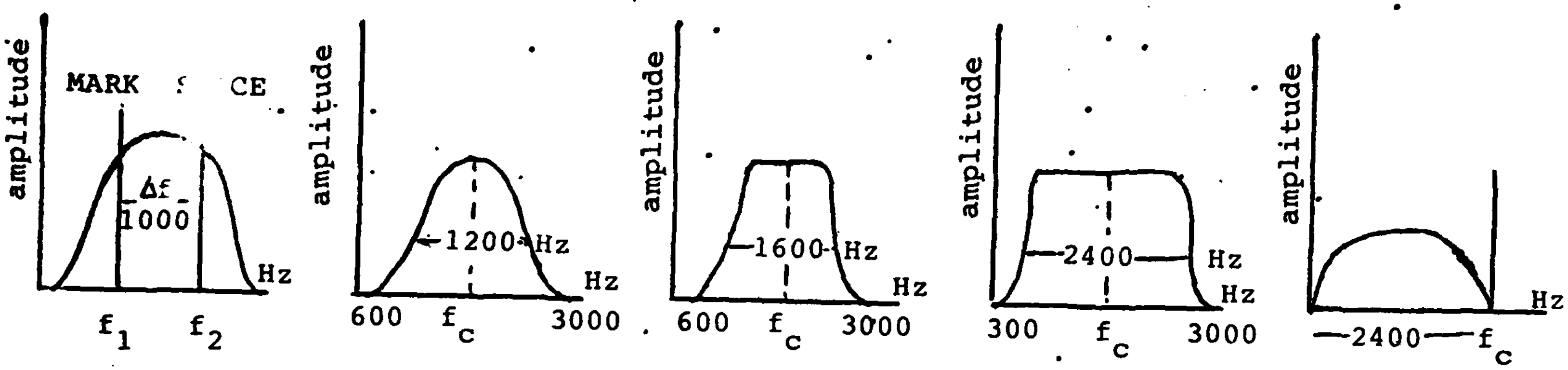


Figure 5.1.2.2 a) modified baseband response and
b)corresponding time response



FSK: f_1=1200Hz f_2=2200Hz	PSK suppressed carrier @ 1800Hz	PSK suppressed carrier @ 1800Hz	QAM suppressed carrier @ 1650Hz	SSB carrier @ 2800Hz
1200 to 1800 bps	1200 baud 4-Phase 2400 bps	1600 baud 8-phase 4800 bps	2400 baud 16 points 9600 bps	4800 baud 4 levels "in" 7 levels "out" 9600 bps

Figure 5.1.2.3 Spectral Densities for five types of modems from Kretzmer [23]

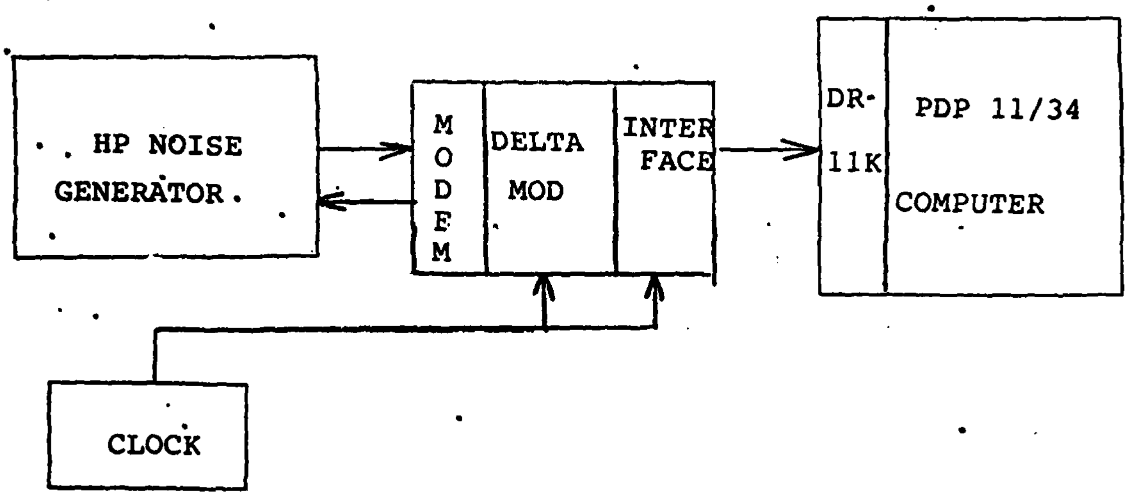
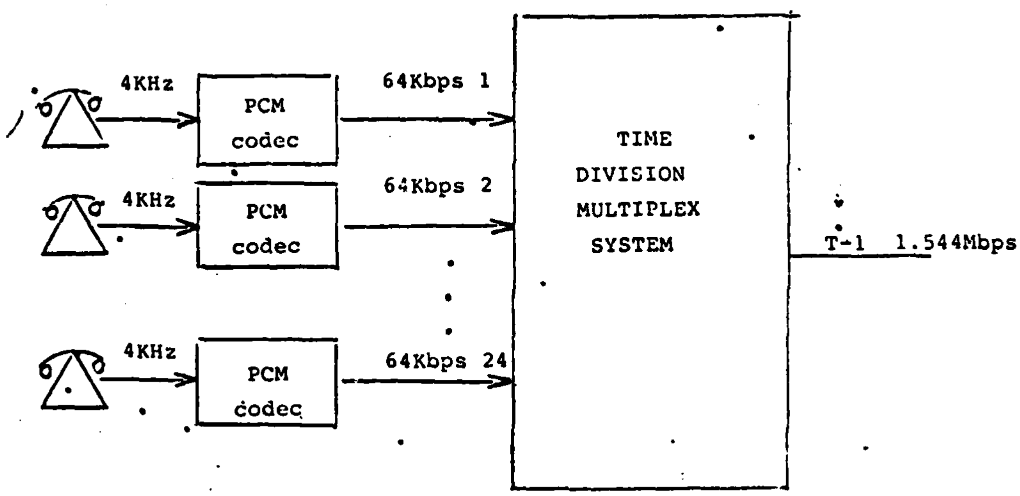
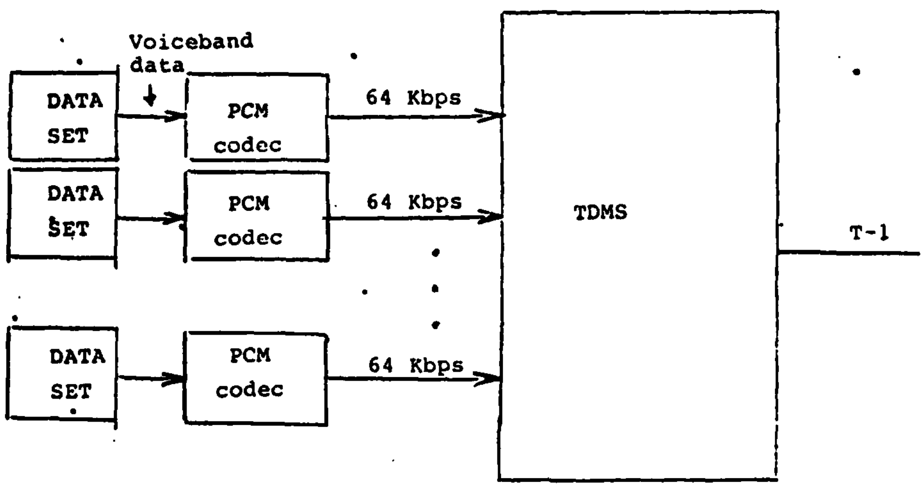


Figure 5.2.1 Block diagram for measuring correlation of voiceband data



a) Voice Concentrator



b) Data Concentrator

Figure 5.3.1 Present strategy for digital transmission

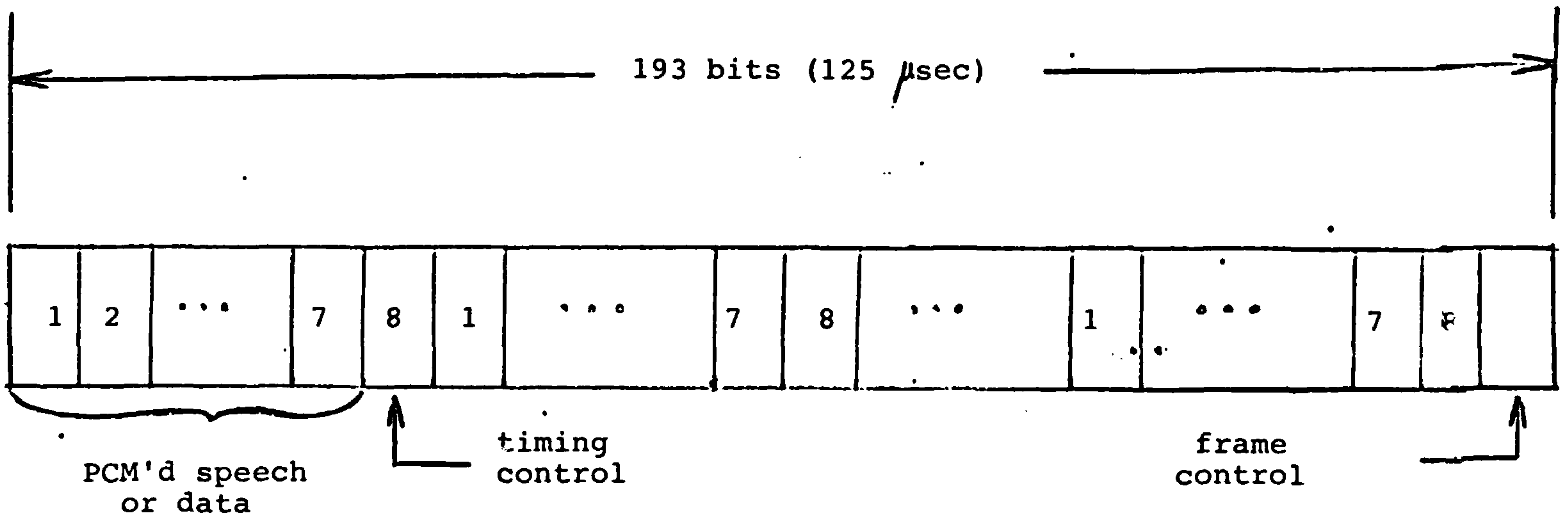


Figure 5.3.2 Format used for digital telephony (1 frame)

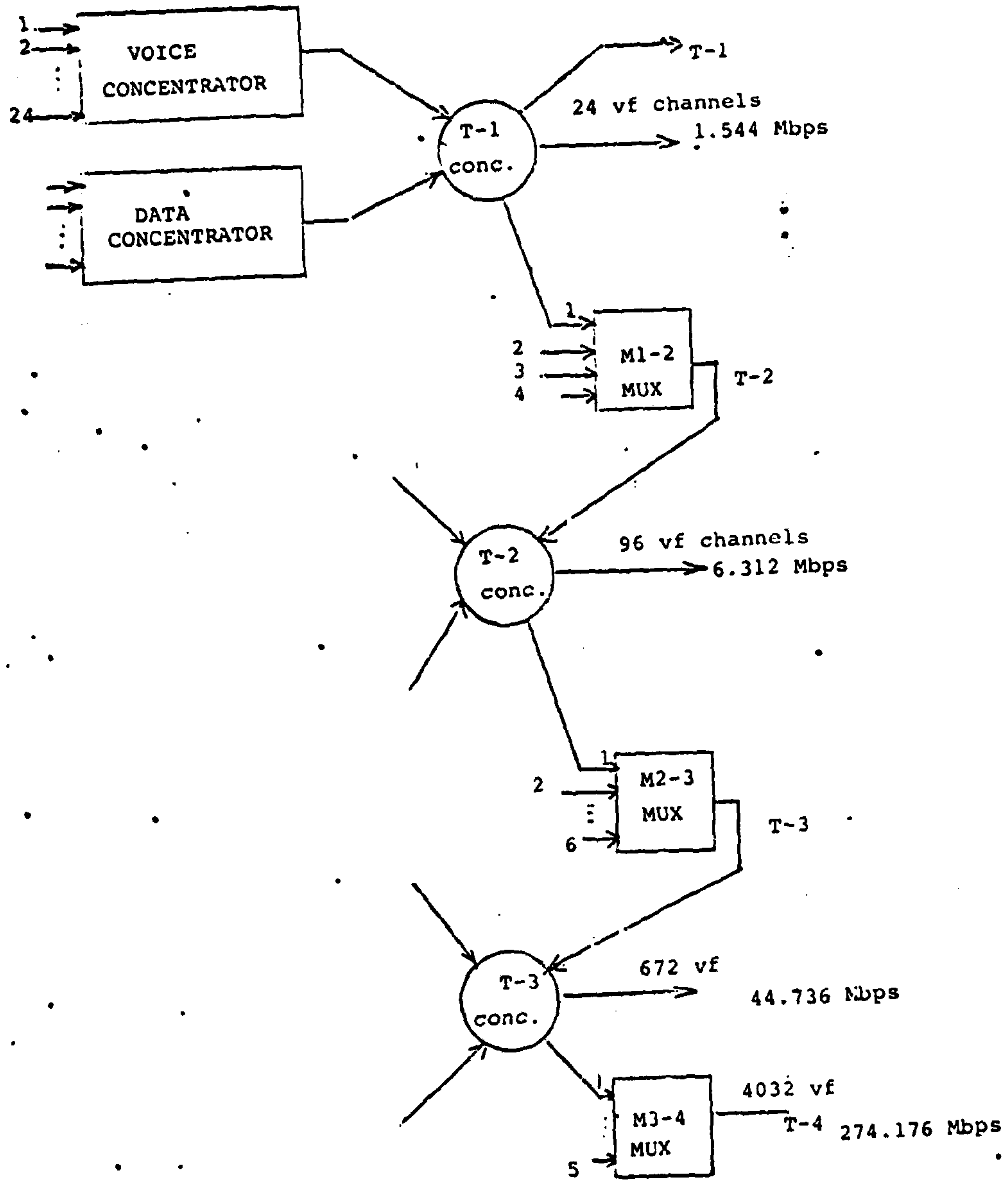


Figure 5.3.3 Hierarchy of digital telephone transmission in the USA

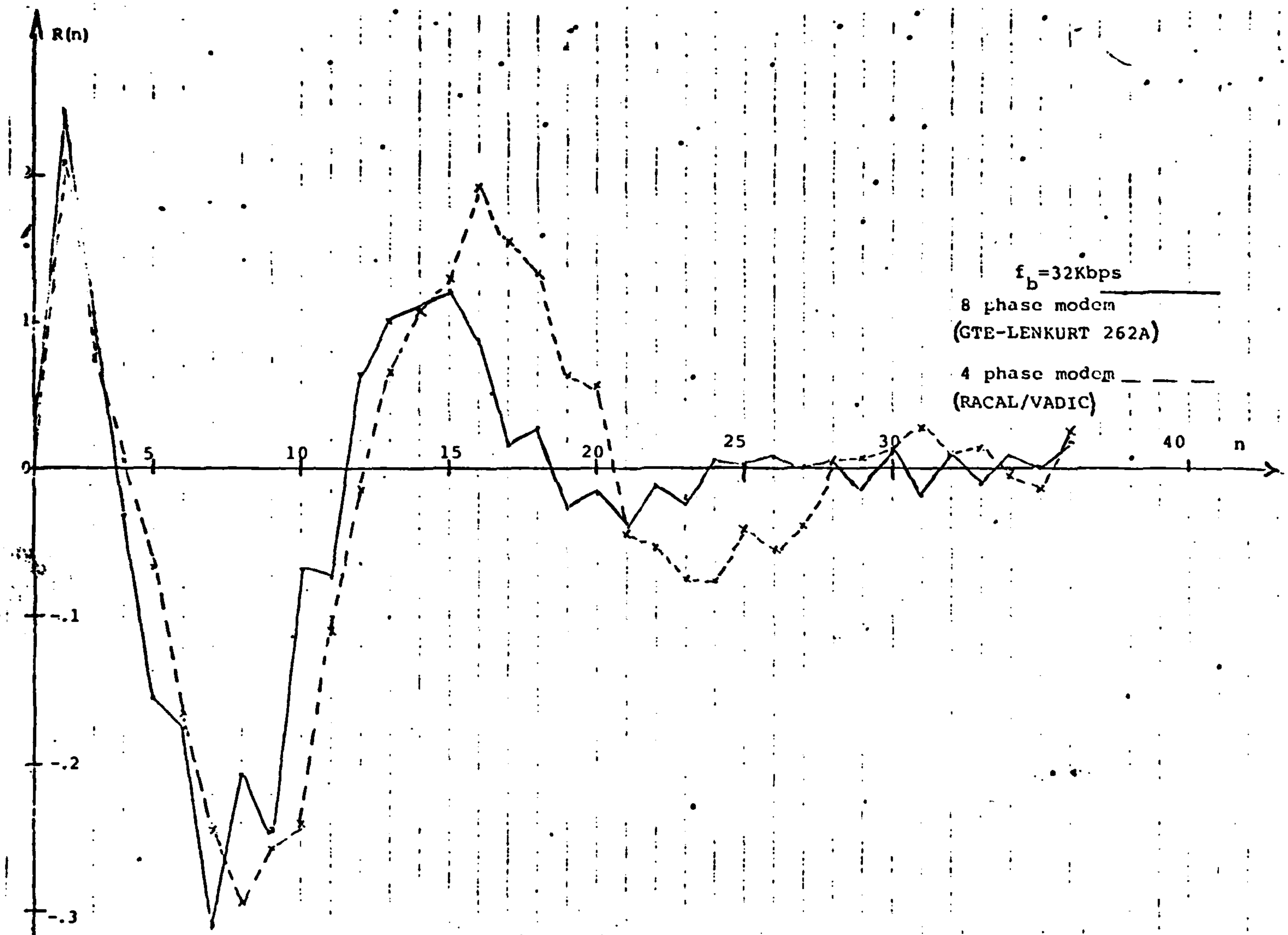


Figure 5.4.1 Autocorrelation of DM digital output sampling 4phase and an 8phase modems

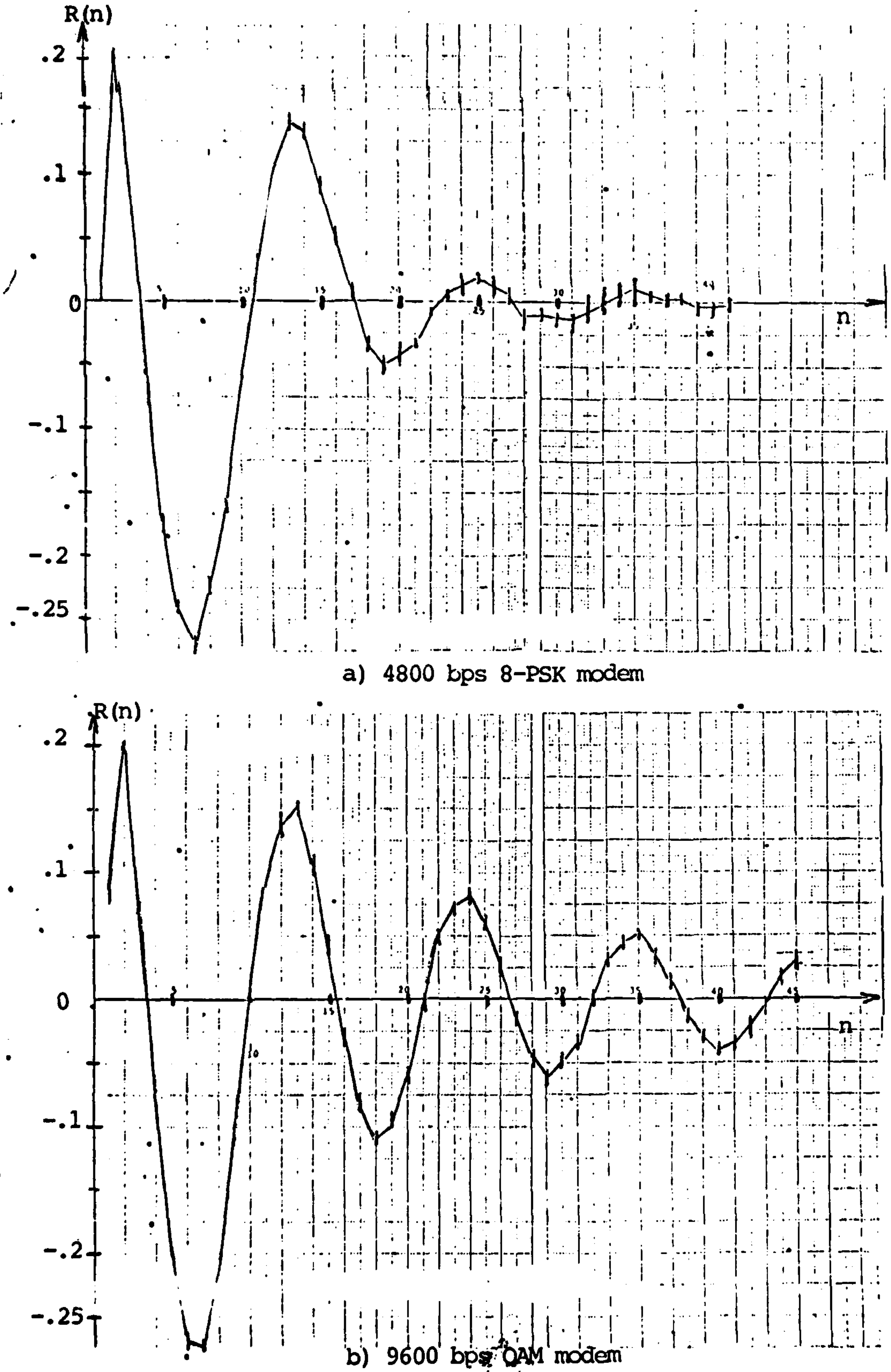
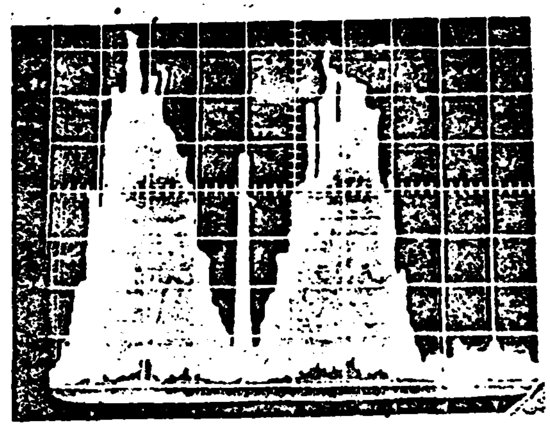


Figure 5.4.2 Autocorrelation function of 4800 bps (a) and 9600 bps (b) modems sampled at 32Kbps using ensemble averaging techniques



Measured spectral density of 4800 bps modem (1Khz/div)

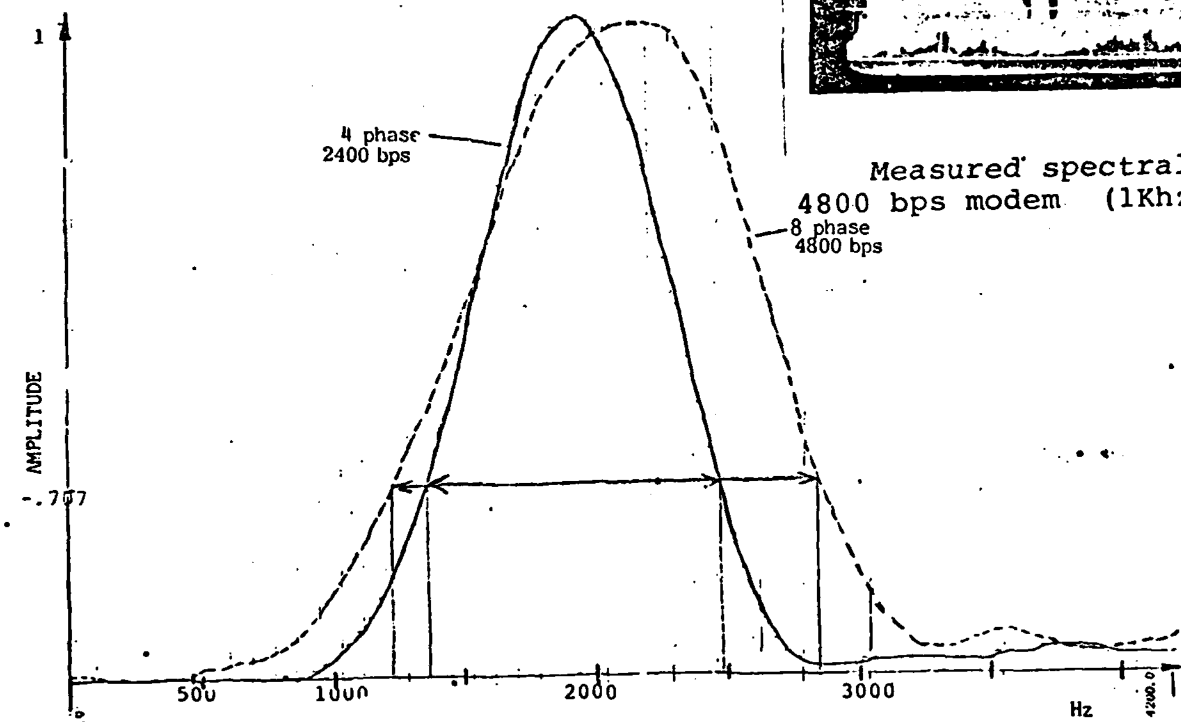
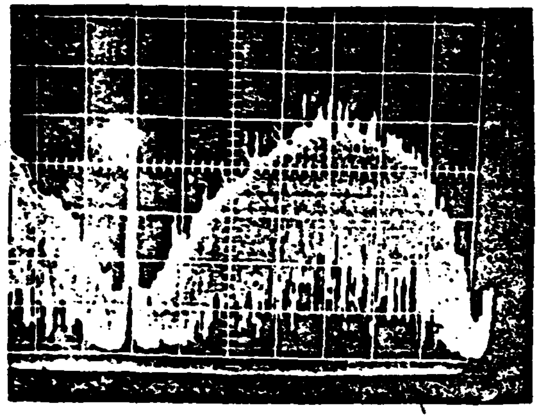


Figure 5.4.3 Spectra density of 2400bps modem and 4800bps modem sampled @ 32Kbps



Measured Power Spectral
Density of 4800 bps
8-PSK modem
(500 Hz/div)

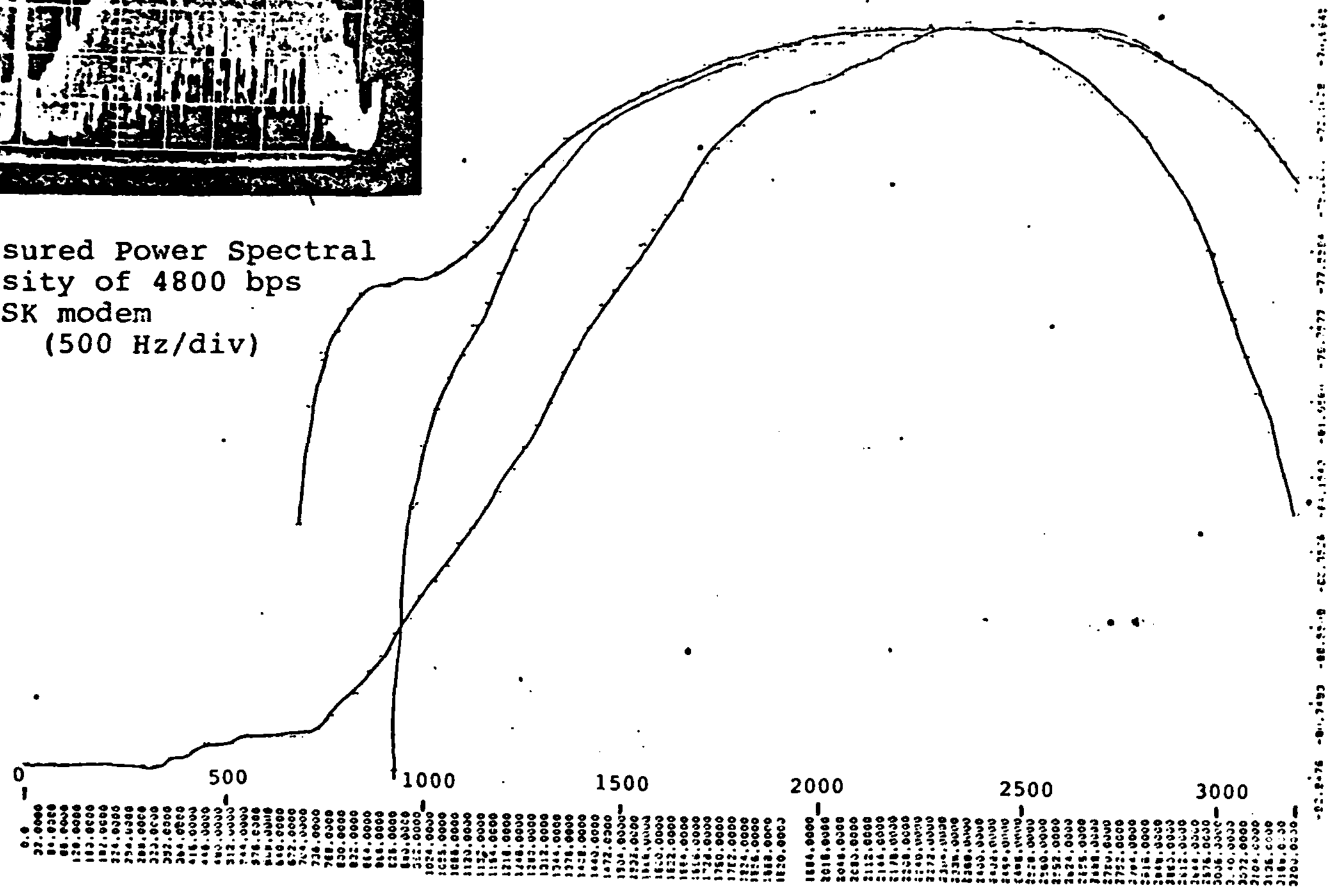
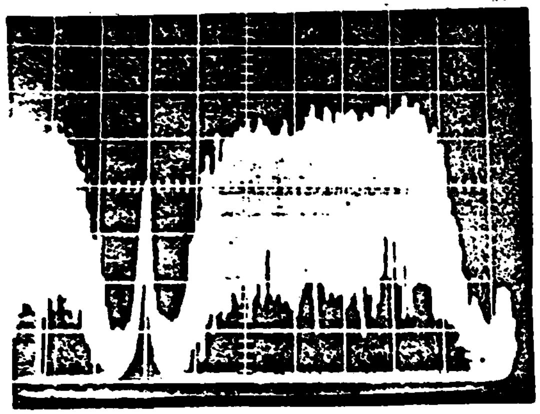


Figure 5.4.4 Calculated PSD of 4800bps 8-PSK modem



Measured Power Spectral
Density of 9600 bps
16-QAM modem
(500 Hz/div)

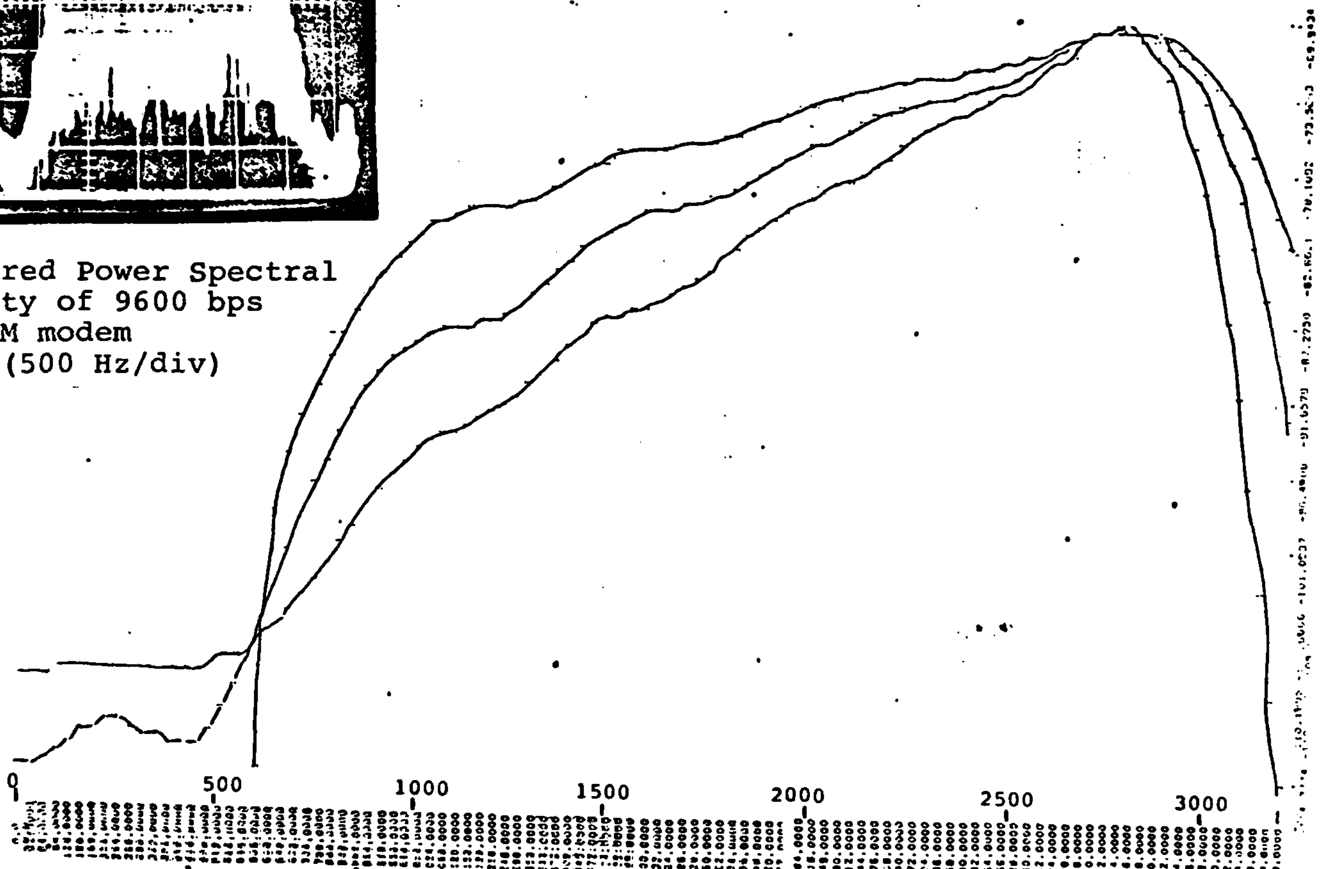


Figure 5.4.5 Calculated PSD of 9600bps QAM modem

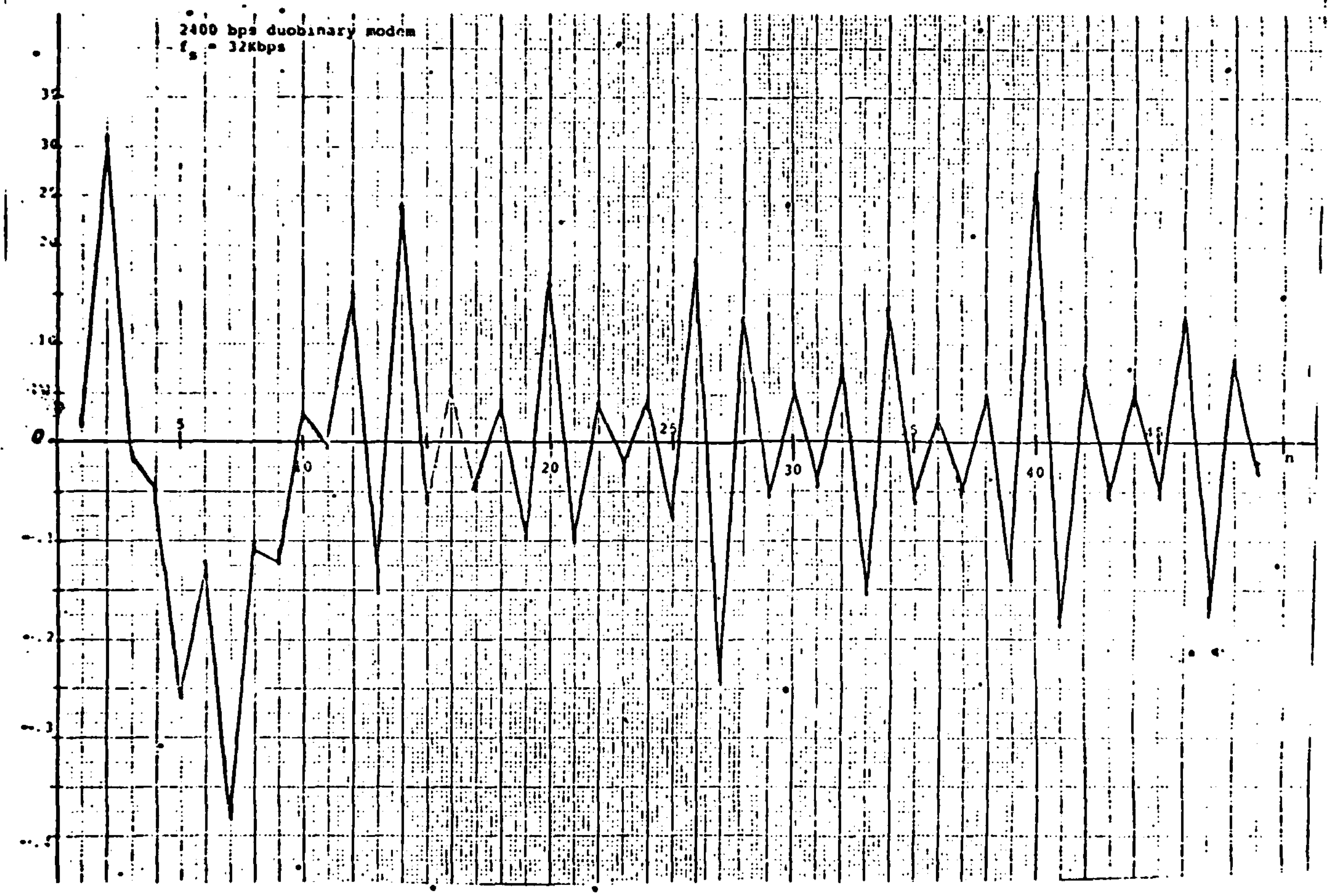
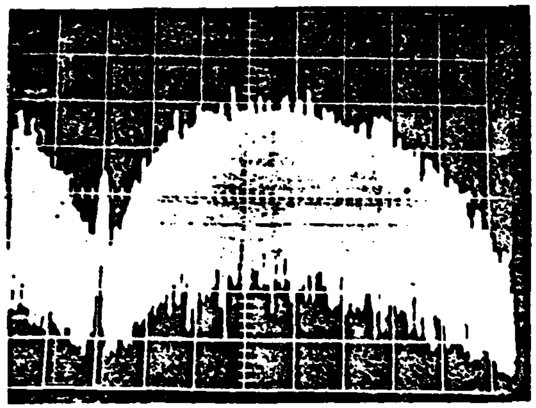


Figure 5.4.6 Measured autocorellation of 2400bps duobinary modem using time averaging techniques



Measured Power Spectral density of 2400 bps duobinary modem (500 Hz/div)

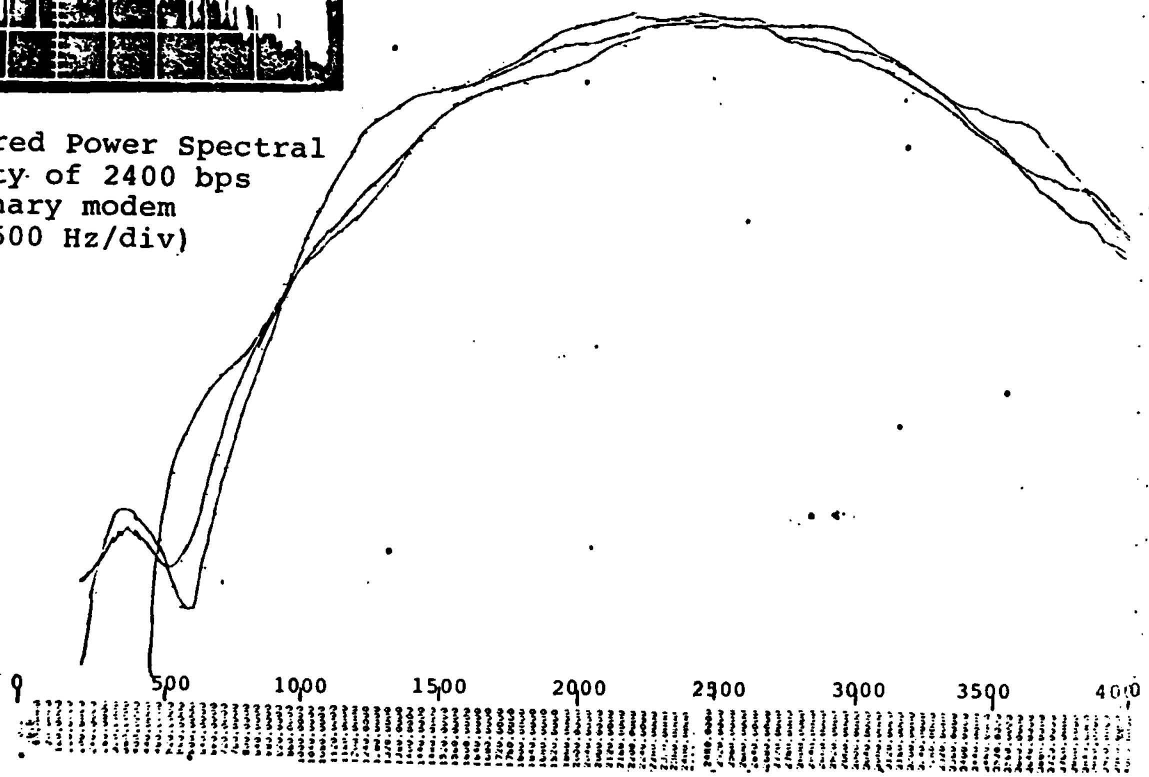
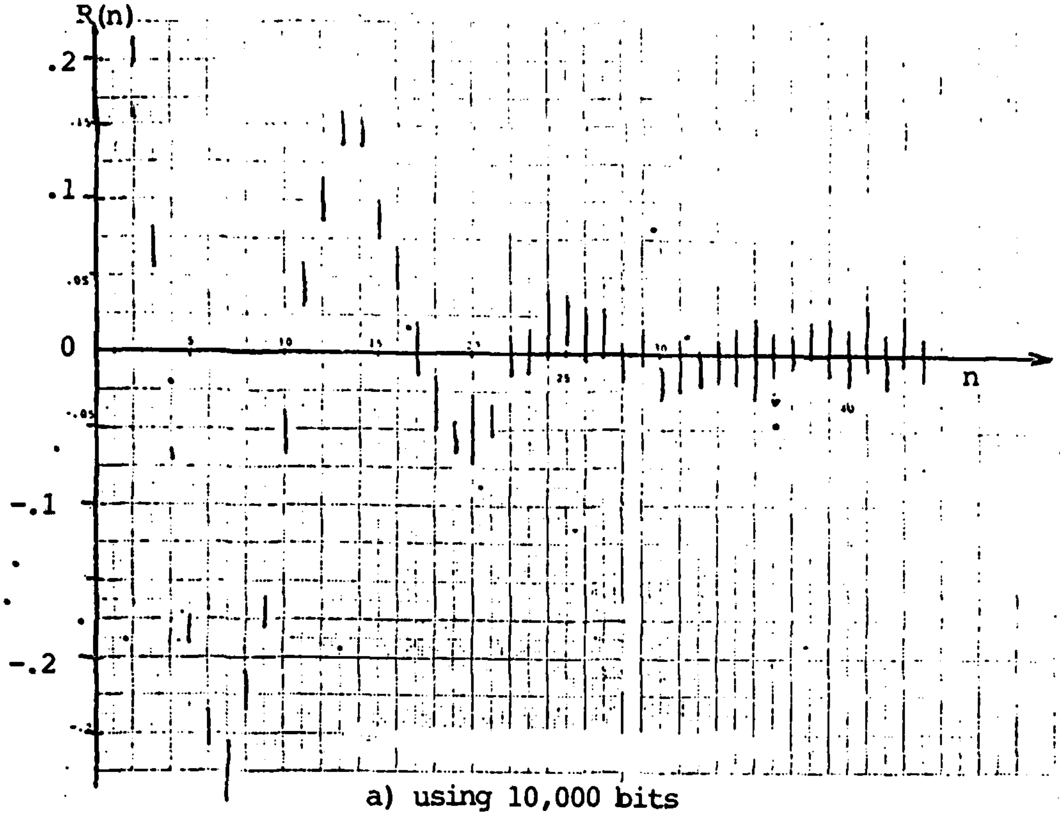
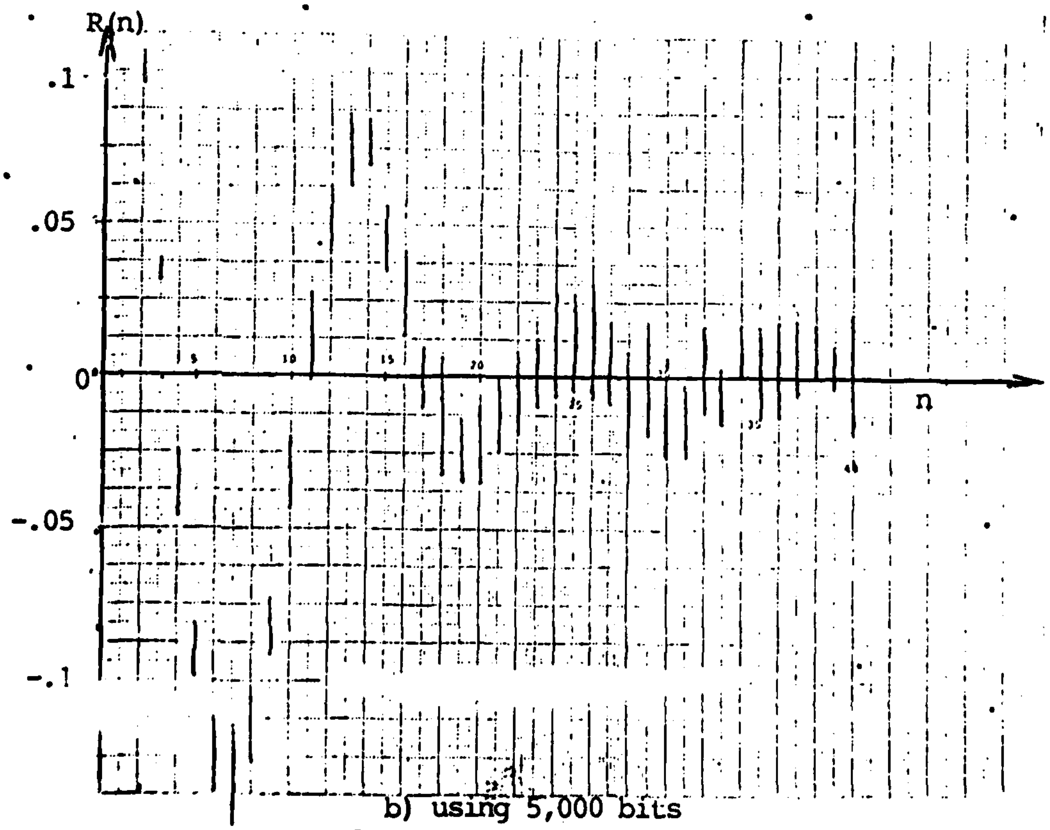


Figure 5.4.7 Calculated PSD of 2400 bps duobinary modem

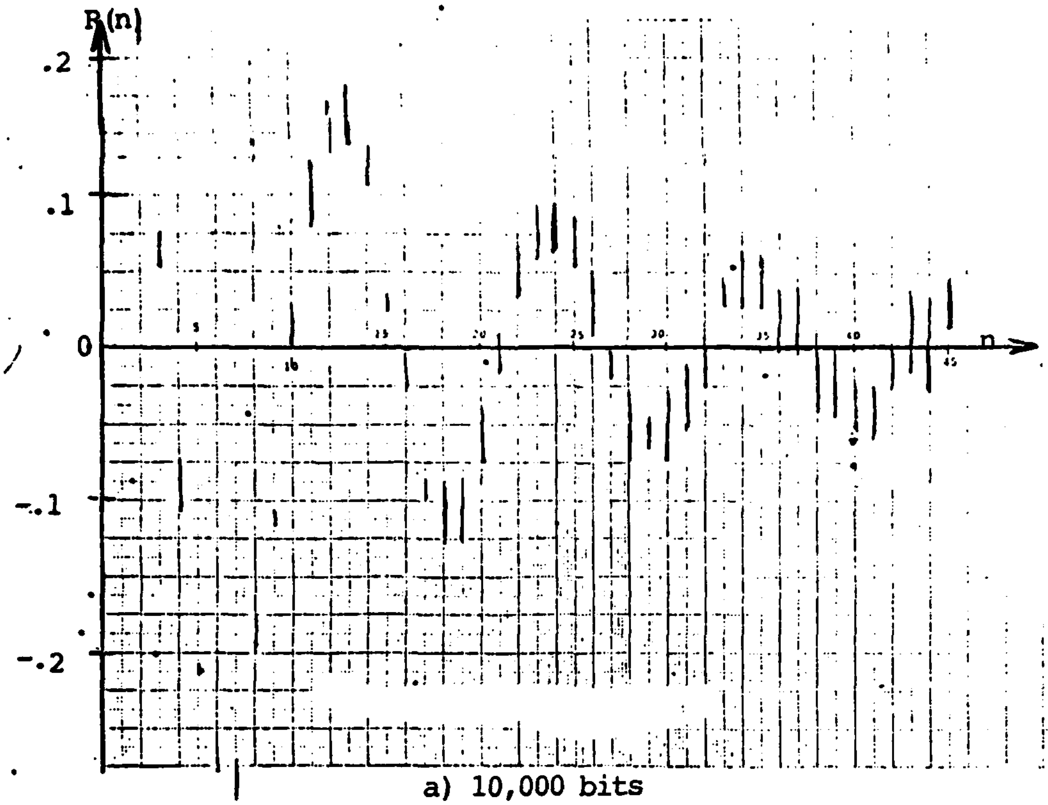


a) using 10,000 bits

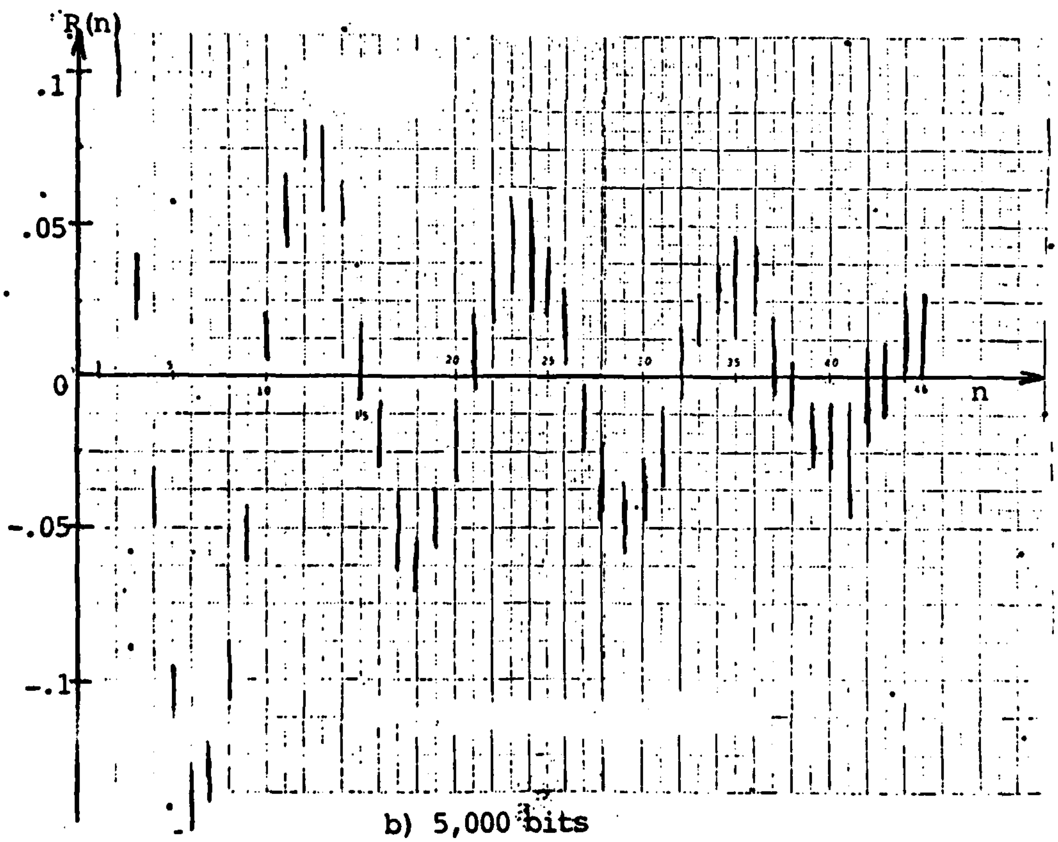


b) using 5,000 bits

Figure 5.4.8 Autocorellation of 4800 bps GTE-Lenkurt modem sampled at 32 Kbps using time averaging techniques with a) 10,000 bits and b) 5,000 bits



a) 10,000 bits



b) 5,000 bits

Figure 5.4.9 Autocorellation of 9600 bps WESTERN ELECTRIC modem sampled at 32 Kbps using time averaging techniques with a) 10,000 bits and b) 5,000 bits

	NO. of TRIALS	NO. of ERRORS	P_{error}
	45,184	23	5.09×10^{-4}
	2,494	0	$<10^{-5}$
	18,089	11	6.08×10^{-4}
	20,000	4	2.00×10^{-4}
	23,000	10	4.34×10^{-4}
Total	108,767	48	4.40×10^{-4}

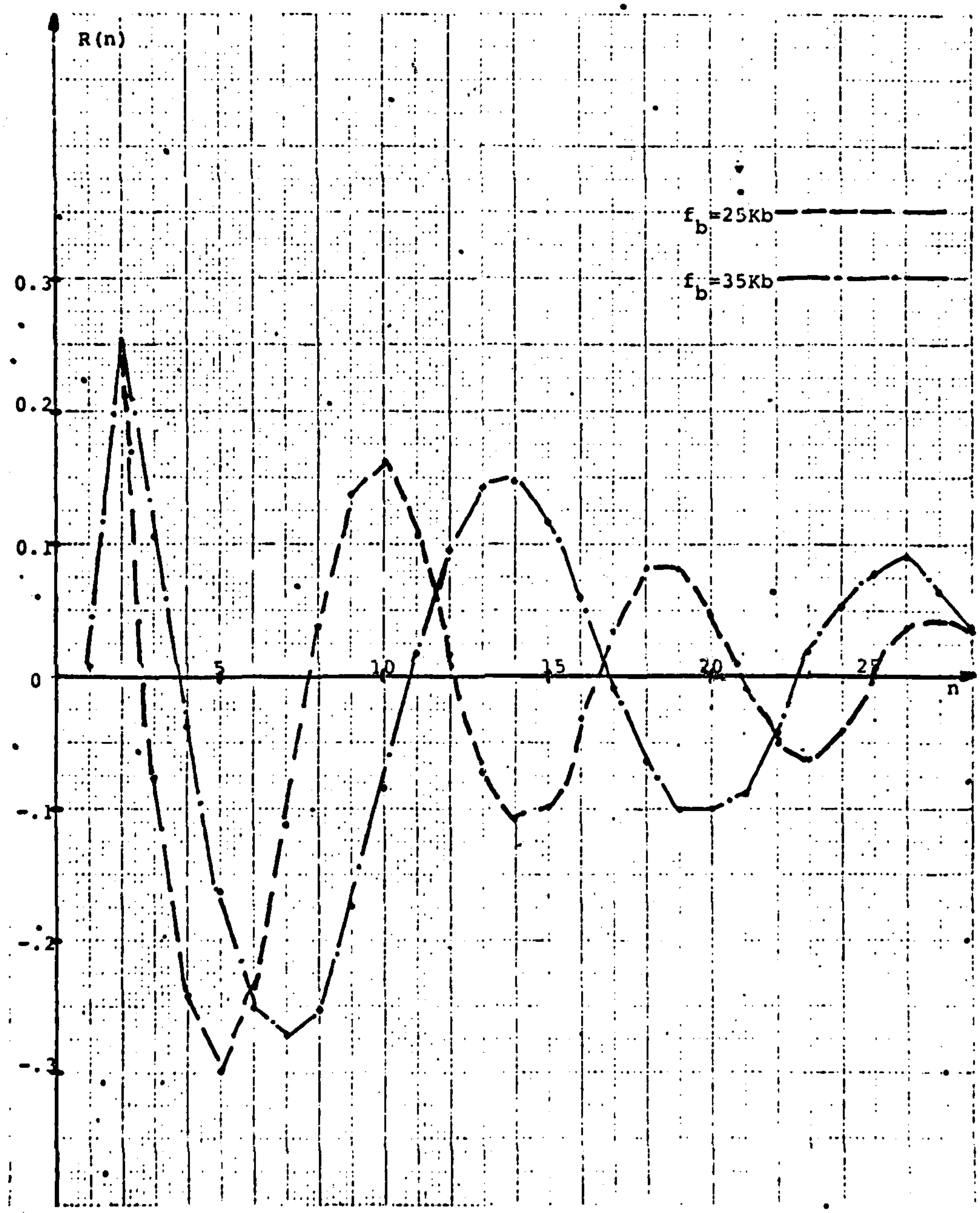
a) threshhold $<$ 2800

	20,000	2	1.00×10^{-4}
	22,080	4	1.80×10^{-4}
	30,000	14	4.67×10^{-4}
	3,776	0	$<10^{-5}$
	27,500	8	2.91×10^{-4}
Total	103,356	28	2.70×10^{-4}

b) threshhold $>$ 2800

$$\text{Threshhold} = (-(n_{17}+n_{18}+n_{19})+(n_{23}+n_{24})-n_{16})$$

Table 5.5.1 Results of Modem discrimination experiments
a) 4800 bps modem b) 9600 bps modem



F e 5.5.1 Autocorrelation of DM output of 9600bps modem for various sampling rates

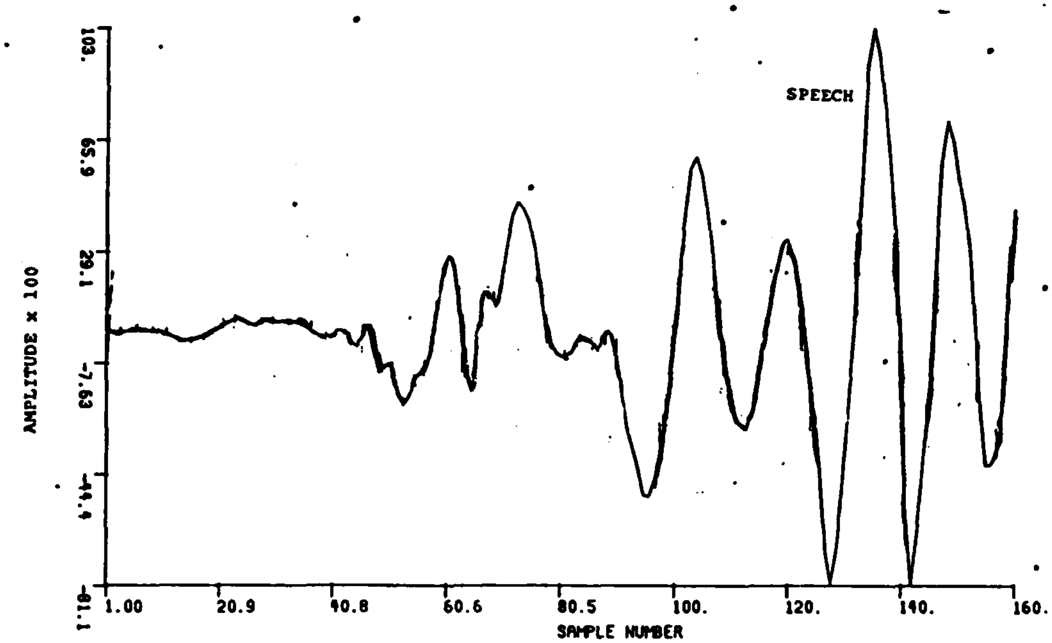


Figure 5.6.1 Waveform of speech signal from O'Neal [21]

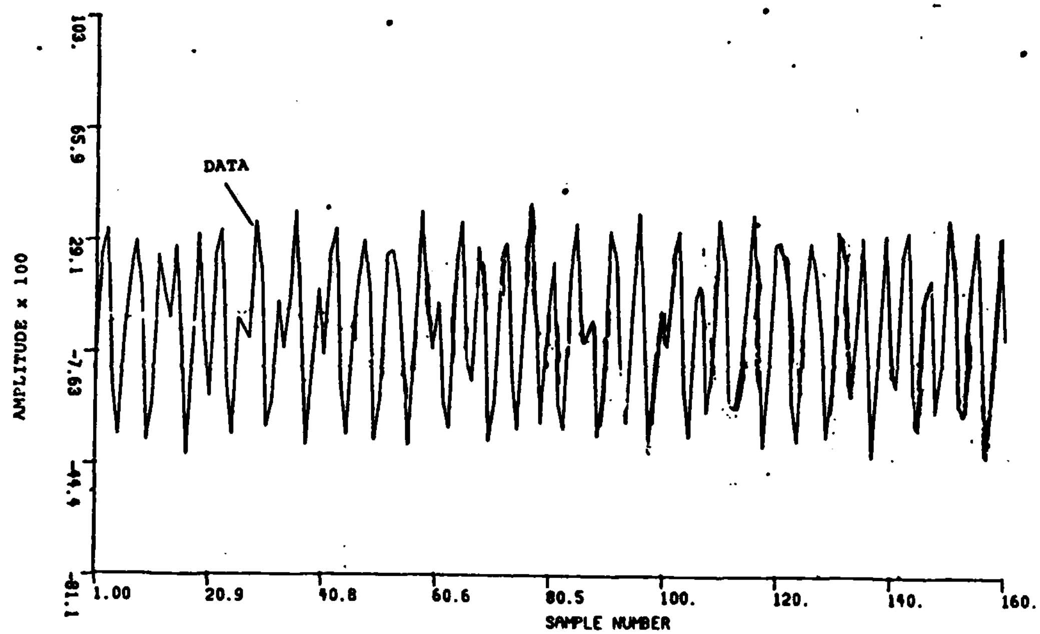


Figure 5.6.2 Waveform of PSK data signalfrom O'Neal [21]

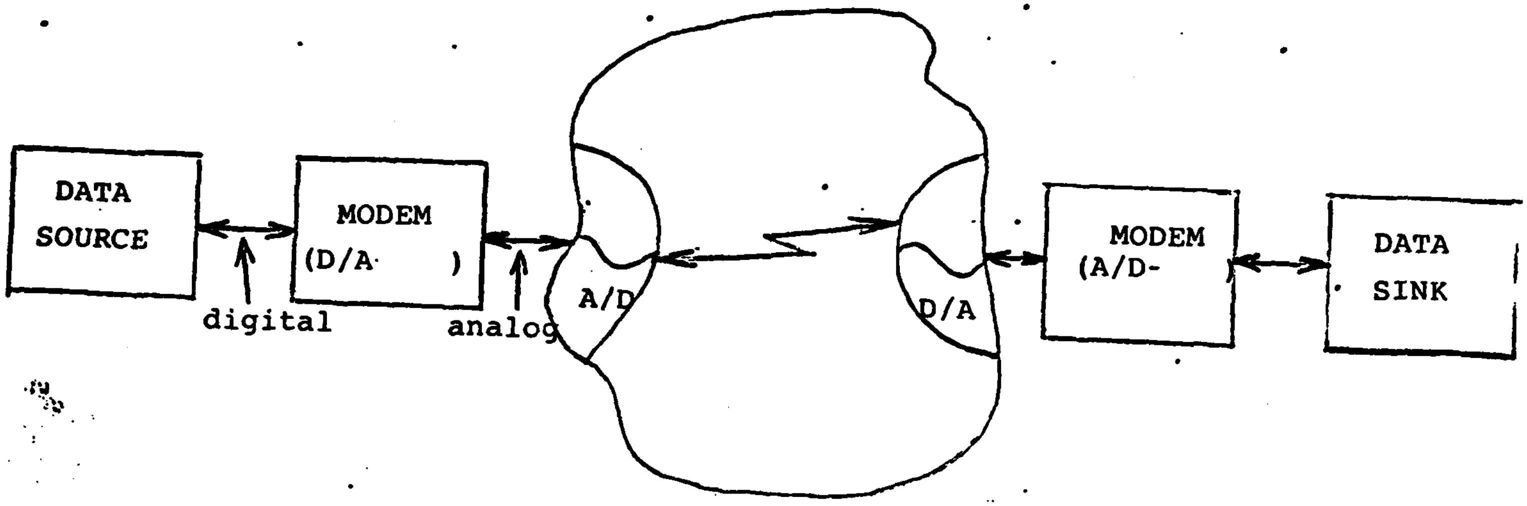


Figure 5.7.1 Format conversion for digital transmission of data signals

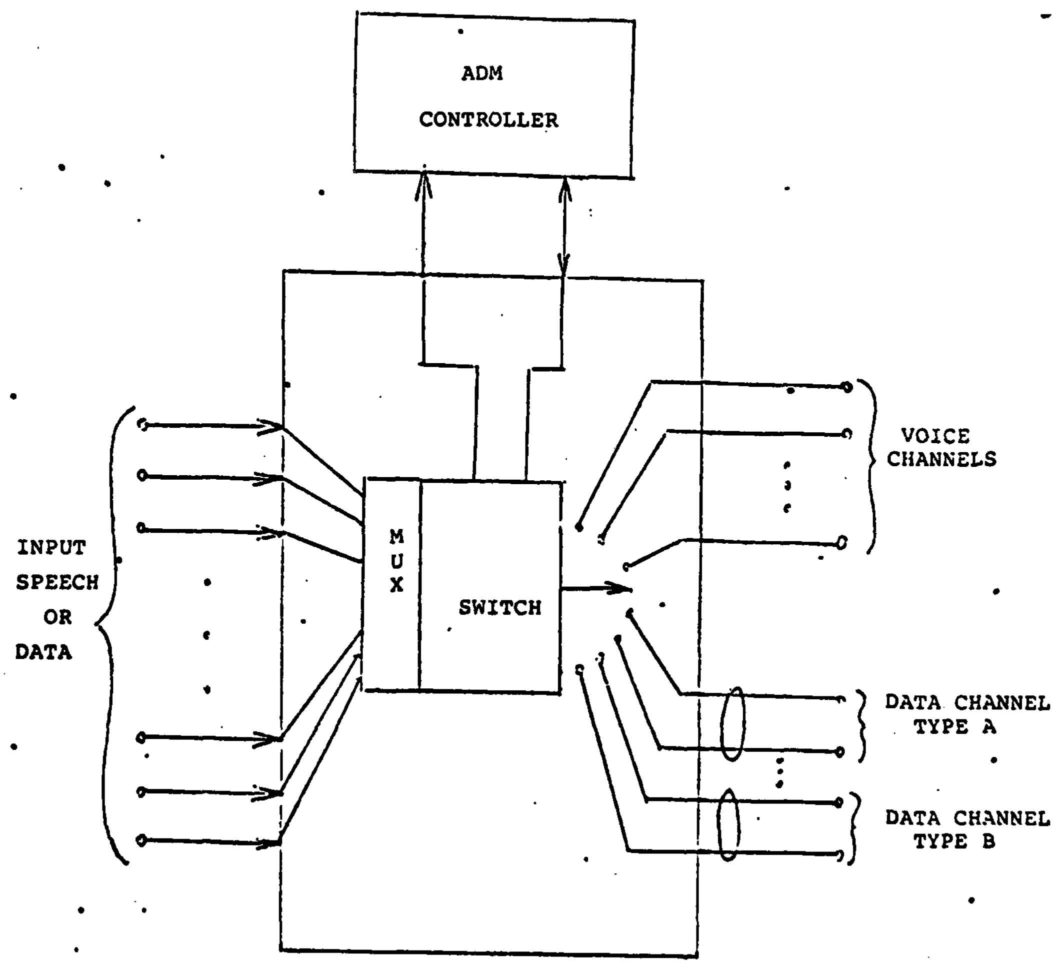


Figure 5.7.1.1 Automatic router

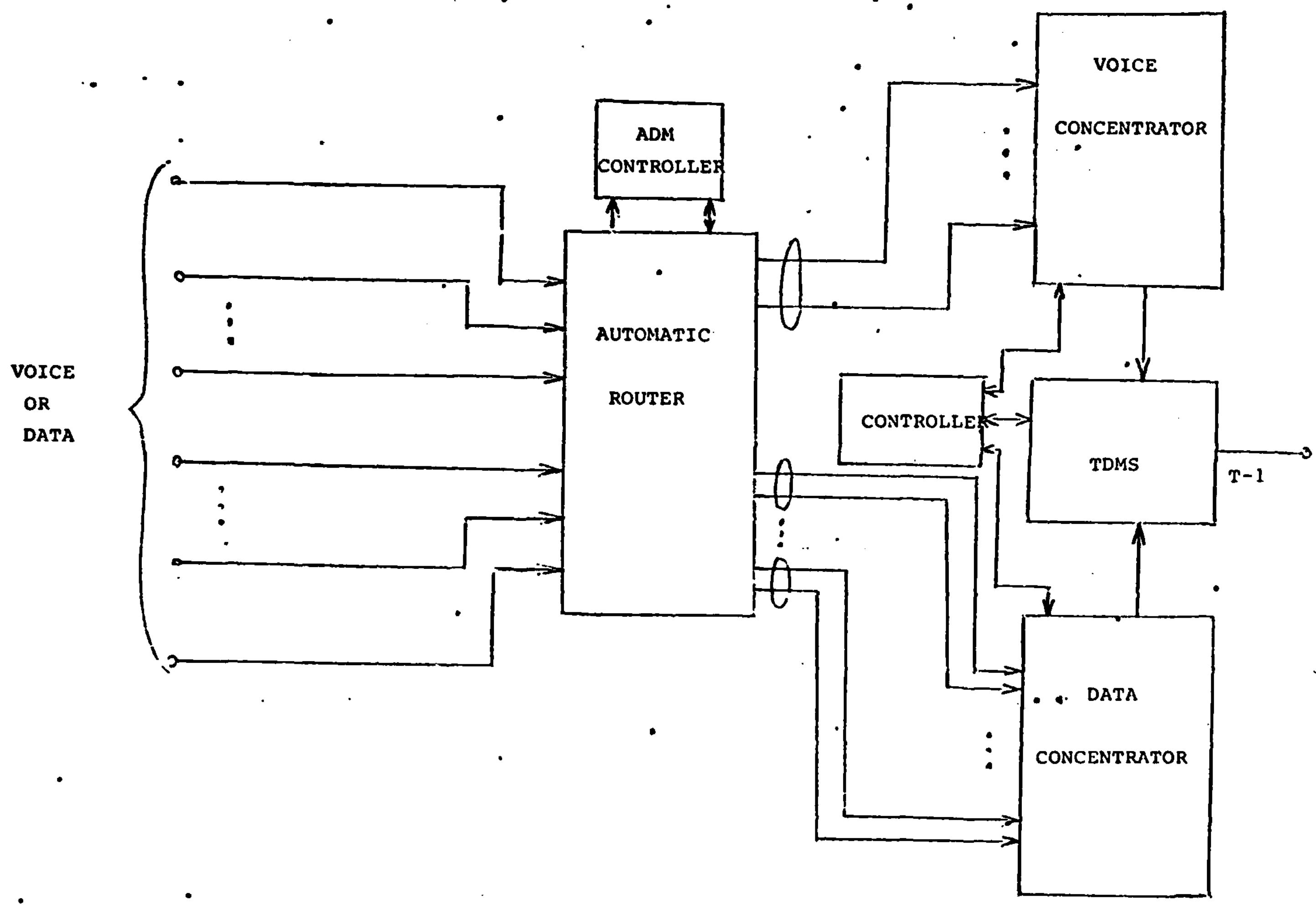


Figure 5.7.2 Proposed Automatic routing telephone system

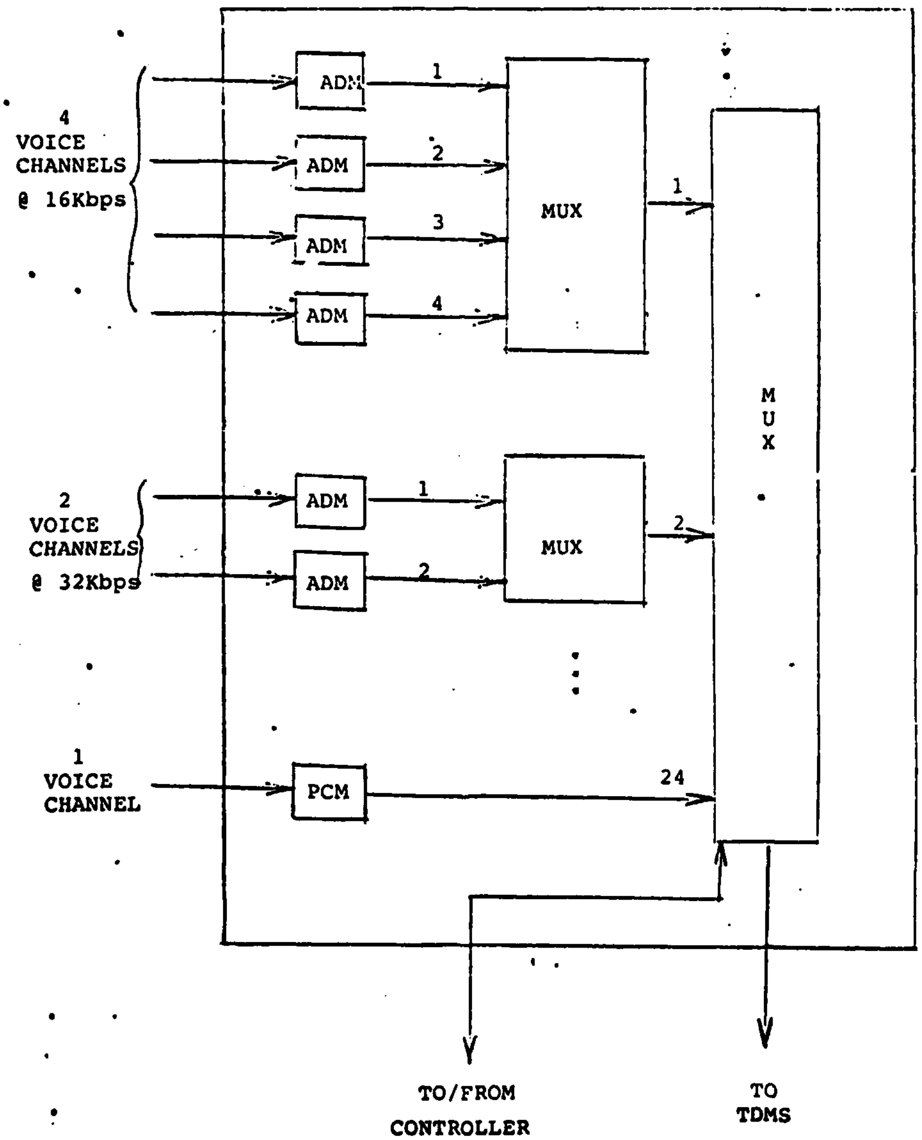


Figure 5.7.2.1 Speech concentrator

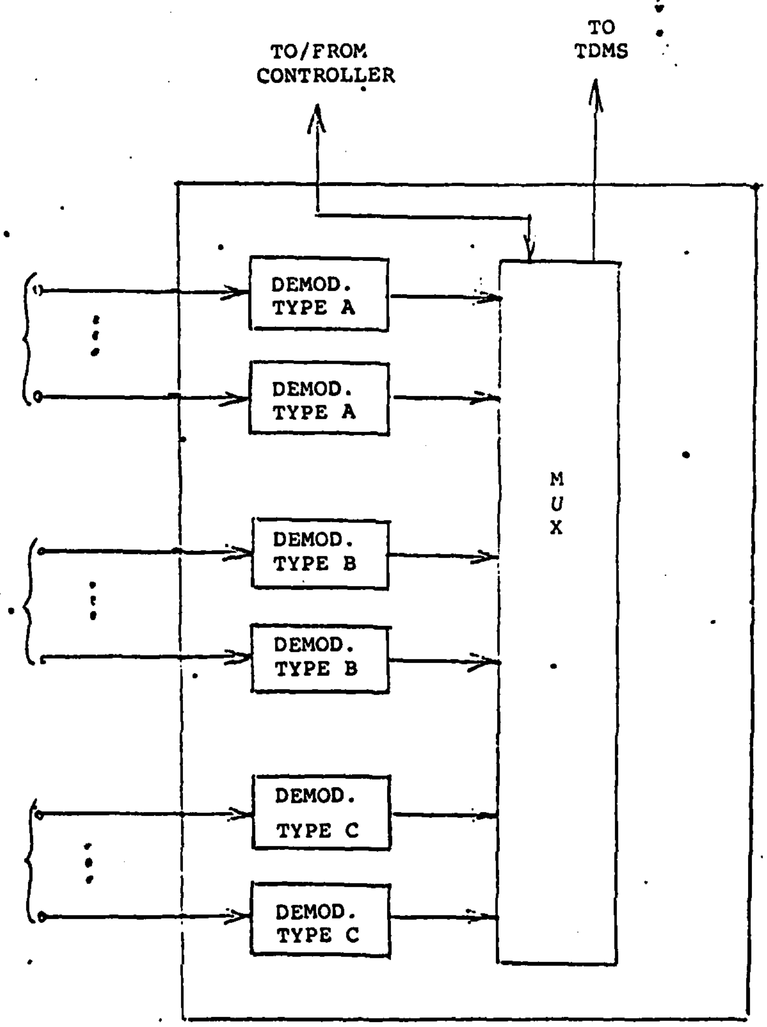


Figure 5.7.3.1 Data concentrator

CHAPTER 6

DISCUSSION AND CONCLUSIONS

This dissertation includes two areas of research. One was the development of a real-time packet voice network simulator, the second was voiceband modem/speech discrimination and identification.

Chapters 1 thru 4 are concerned with the first part of the dissertation. In chapter 1 the packetization process is introduced. A discussion, as to the amenability, of speech for inclusion into the packetization environment is presented.

In chapters 2 and 3 the specifics of a real-time packet voice simulator are presented. In these chapters the software (e.g. the philosophy of operation) and external hardware connection are discussed. The importance of the silence/speech detection schemes was followed by a discussion of two specific algorithms (using the ADM) to exercise the packet voice simulator. For the simulator the Song Voice Adaptive Delta Mod (SVADM) was chosen as the encoding scheme. This algorithm was used in this research for two reasons. The first was the availability of th DM hardware and the ability to modify the encoder/decoder pair. The second, was due to past work which showed that the SVADM was the preferred encoding scheme at 16Kbps

sampling rate. It should be noted that various other types of encoding schemes can easily be used with the simulator.

Chapter 4 presents the results of tests performed on the simulator with DM encoding of the input for the two algorithms discussed in chapter 3. It is seen that the packet size statistics for the two algorithms differ slightly. The word-by-word s/s algorithm has a higher (as much as 10%) rate of maximum packet transmitted (greater than 90%) than the 16 bit s/s algorithm. Packets ranging in size from 1 thru 127 bytes are approximately uniformly distributed with values less than 2.5% of the total transmitted packets. Total packet transmission rate statistics and quality curves show that at packet transmitted rates of 75% (of maximum) there is good quality of the transmission.

In chapter 5 the second part of the dissertation is presented. It is seen that various modems exhibit different autocorrelation functions (or equivalently spectral density). The autocorrelation function measured was of the digital output of the DM. This was done with the aim of distinguishing various modems, one from the other. It is necessary that the process of modem identification be done at a fast a rate as possible with probability of error less than 10^{-5}. Tests were performed on various modems at DM sampling rate of 32Kbps. Error rates of approximately 10^{-4} were measured. The process per

each trial was reduced to 250-500msec using time averaging techniques. It is possible at higher sampling rates to reduce the probability of error and the time involved in the measurement.

The aim of the first part of the research presented in this dissertation was to develop a real-time packet voice network simulator, as outlined in section 1.5. Statistical results of two schemes for silence/speech detection, using ADM encoding, were presented. It was shown that with efficient s/s etection schemes, the percentage of maximum size packets transmitted can exceed values of 90% of the total packets transmitted. This compares with a previous study [5], where only 50% of the transmitted packets were of maximum size. The simulator also allowed for the study of a two way conversation. The random number generator was used to examine the physiological effects on a conversation caused by packet loss and random delay experienced by the packets. This is of importance in establishing values for maximum allowable delay for buffering the received voice packets at the receiver.

In the second part of the dissertation it was shown that the ADM can be used to discriminate various voiceband modems; one from the other. It is concluded from the resu presented in chapter 5 that various modems can be distiguished one from the other at error rates less than 10^{-5} and at times much faster than that of the study, which

was due to the limitations of the computer used. This is encouraging since a study performed by Yatsuzuka [14], using signal energy and zero crossing rates to distinguish various modems did not achieve good results

6.1 SUGGESTIONS FOR FUTURE WORK

Further research could easily be applied as a result of this disseration. The program for the packet voice simulator can easily be adapted to do studies in packet voice/data integration and development of teleconferencing protocols, all in real time. Various other encoding schemes can easily be applied and studied using the versatility of the simulator. Other applications of the simulator can be found in local area networking as well as large internetwork packet systems. Voiceband modem/speech waveform identification and discrimination can find applications in local area networks and over large nets where the modems used can be restricted.

REFERENCES

1. D. Cohen, "Specifications for the Network Voice Protocol", NSC note 68 (RFC471, NIC 42444), Jan. 1976.

2. S.L. Cansey, E.R. Masler, E.R. Cole, "Some initial measurements of the ARPANET packet voice transmission", Conf. Rec. NTC 1978, Birmingham, Ala. pp.12.2.1-12.2.5, Dec.1978.

3. V.R. Dhadesugoor, C. Ziegler, D.L. Schilling, "Delta Modulation in packet voice systems", IEEE Trans. Comm. vol. comm-28, Nov 1979.

4. C.L. Song, J. Garodnick, D.L. Schilling, "A variable step size robust Deltamodulator", IEEE Trans. comm. Tech. vol. com-19, Dec. 1971.

5. V.R. Dhadesugoor, "Delta Modulation systems for voice communications", PhD dissertation Graduate School of CUNY NY,NY Jan 1981.

6. L. Kleinrock, "Principles and lessons in packet communications", Proc. of IEEE vol. 66 Nov. 1978, pp.1320-1330.

7. D.W. Davies et al, "Computer Networks and their protocols", John Wiley and Sons, N.Y. 1979.

8. J. Gruber," Delay in integrated voice and data networks", IEEE trans. Comm. vol. comm-29, June 1981.

9. G Coviello, O.L. Lake, G.R. Redinbo, "system design implications of packetized voice", ICC conf. rec. Chicago, Illinois June 1976.

10. G.J. Coveillo, " Comparative discussion of circuir vs packet switched voice", IEEE trans comm., vol comm-27 pp 1153-1160, Aug. 1979.

11. P.T. Brady, " A statistical analysis of on-off patterns in 16 conversations", BSTJ pp.73-91, June 1968.

12. S.T. Campanella, "Digital Speech Interpolation",Comsat Tech. Rev., vol. 6 pp. 127-158 spring 1976.

13. R.V. Cox and R.E. Crochiere, " Multiple user variable rate coding for TASI and packet transmission system", IEEE Trans. Comm. vol comm-28 pp.334-337 mar. 1980.

14. Y. Yatsuzuka, " High gain DSI-ADPCM", ICASSP rec. Wash. DC Apr. 1979 pp.436-441.

15. J.M.Elder, J.F. O'Neill " A speech interpolation system for private networks", NTC rec. Vol. 1 Birmingham Ala. 1978 pp.14.6.1-14.6.5

16. T.R. McPherson Jr., J.B. O'Neal, "PCM speech compression via ADPCM-TASI", ICASSP rec. 1977 pp.184-187.

17. J.W. Forgie, " Speech transmission in packet switched store-and-forward networks", Proc. NTC 1975 pp.137-142..

18. N.S. Jayant, " On the power spectrum of the staircase function in linear delta modulation", IEEE Trans. Acoust., speech, sig. processing April 1970 pp. 160-168.

19. W.R. Bennett, " Statistics of regenerative digital trans[illegible]ssion", BSTJ Nov. 1958 pp. 1501-1542.

20. J. Martin, " telecommunication and the computer", 2nd ed., Prentice-Hall, Engelwood cliffs N.J., 1976

21. J.B. O'Neal Jr., " Waveform encoding of voiceband modem waveforms",Proc. of IEEE, Feb. 1980 pp. 232-247

22. R.W. Lucky, J.Salz, E.J. Weldon Jr., " Principles of Data Communications", McGraw-Hill, New York 1968

23. E.R. Kretzmer, " The evolution of techniques of Data Communications over voiceband channels"; IEEE comm. soc. mag., Jan. 1978 pp. 10-14

24. R.W. Lucky, P.E. Green Jr., " Computer Communications", IEEE Press, New York 1975

25. A. Lender, " The duobinary technique for high speed data transmission", IEEE winter conference, 1963, conf. paper cp63-283

26. D. Minoli, " Optimal packet length for packet voice communications", IEEE Trans. comm. Com-27, 1979 pp.607-611

27. R.W. Hamming, " Numerical methods for Scientists and Engineers", McGraw-Hill, N.Y. 1962

END

FILMED

10-83

DTIC